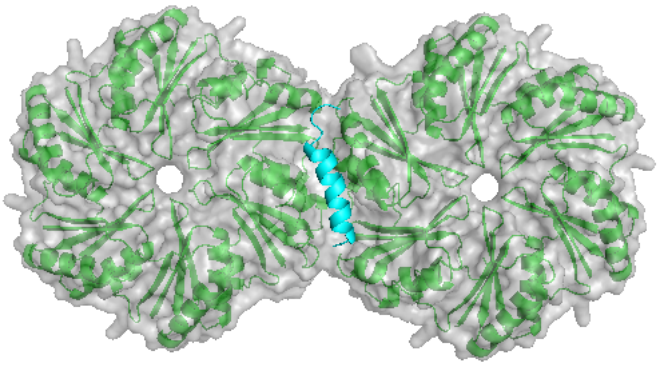

DOCTORAL THESIS

Elucidating shell protein-enzyme interactions within the bacterial propanediol utilisation microcompartment

Author: Jack BRADLEY-CLARKE

Supervisor: Richard PICKERSGILL

A thesis submitted in fulfillment of the requirements

for the degree of Doctor of Philosophy

Pickersgill Lab

School of Biological and Chemical Sciences

February 2, 2021

Declaration of Authorship

I, Jack BRADLEY-CLARKE, declare that this thesis titled, "Elucidating shell protein-enzyme interactions within the bacterial propanediol utilisation microcompartment" and the work presented in it are my own. I confirm that:

- I, Jack BRADLEY-CLARKE, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below.
- I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material.
- I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.
- I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.
- The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without the prior written consent of the author.
- Details of collaborations and publications:
  Collaborations with the Universities of Kent and Bristol through the BBSRC LOLA grant. X- ray diffraction data presented in chapter 4 was supplied by the University of Kent.

Signed:

_____

Date:

_____

Abstract

Jack BRADLEY-CLARKE

Elucidating shell protein-enzyme interactions within the bacterial propanediol utilisation microcompartment

Bacterial Microcompartments are entirely proteinaceous organelles found in a variety of bacteria. These organelles are composed of shell proteins with five or six- fold symmetry, which confer the compartment with a polyhedral morphology. Bacterial Microcompartments encapsulate either metabolic or catabolic enzymatic pathways, functioning to sequester the substrates of these enzymes within the microcompartment. These enzymes interact with shell proteins through an N- terminal signalling sequence, although precisely how has been largely unclear, with several pieces of work implicating the C- terminus of PduA as being involved. In this thesis, questions regarding how Microcompartments form and how the internalised enzymes are associated with the shell proteins are addressed. Assembly of the shell protein PduA into small oligomers was observed, analysis of these oligomers using MALDI-TOF and native PAGE revealed that they are trimers of PduA hexamers which preclude the formation of higher order structures. Knowledge of these small PduA oligomers was used to assess the interaction between PduA and the encapsulated enzyme PduL. Non- tessellating mutants of PduA were found to not interact with PduL, as assessed by Native PAGE and Circular Dichroism. A combination of computational modelling and site directed mutagenesis demonstrated that the extension peptide on the N- terminus of PduL binds to the interface present where two PduA hexamers interact. The crystal structure of PduL was solved to 1.8 Å and this was then used to model how the whole enzyme might interact with PduA within the facets of the microcompartment shell. This work resolves the seemingly paradoxical

observations that the C- terminus of PduA is present on the exterior of the microcompartment yet is apparently involved in binding enzymes known to be on the interior. The cleft reported here to be the binding site is present on the interior of the microcompartment and is therefore compatible with current knowledge of the orientation of shell proteins in the microcompartment.

Acknowledgements

Thanks to Professor Richard Pickersgill for the opportunity to work on the micro-compartment project, His attention and expertise has proven invaluable throughout the duration of the project and it would not have been feasible without his input.

All members of the G.35 lab have been incredibly supportive and have provided a truly enjoyable, yet productive working environment. Thanks to Ismail, Ben, Yumi, Tania, Jimena, Hui and Shuang who have all been excellent co- workers and friends. In particular I would like to thank Ben for his input and joviality when things were not looking as bright as they could be. Both Yumi and Shuang have been instrumental in teaching me the practical skills required to complete this project and my sincere gratitude goes out to both.

I also extend many thanks to Dr. Ruth Rose, whose advice, input and general assistance in both the practical and personal aspects of this project have been second to none.

Thanks to Roberto in the QMUL mass spectrometry laboratory for his input and in particular his patience when teaching me how to use the MALDI- TOF instrument.

To everyone involved in the BBSRC LOLA programme, our collaborators at the Universities of Kent and Bristol I would like to extend my gratitude. Our tri- annual meetings have been incredibly valuable in assessing the state of the field and the constructive criticism from those involved has been crucial to assessing and improving my own work.

Thank you to the BBSRC and the LIDO programme, The events and support provided by LIDO have been incredibly useful in building skills and confidence through interactions with my peers. In particular a special thanks to Nadine Mogford, the programme organiser.

Finally I sincerely thank my family for their undying support over these four years, and also to my fiance Jade, who has been of immense support and who I would not have met had I not come to London for this project.

Contents

List of Figures

List of Tables

Abbreviations:

1,2- PD	1,2 **P***ropane***d***iol*
2x YT	2x **Y***east and* **t***ryptone*
α	**A**lpha
Adh	**A**lcohol **d***e***h***ydrogenase*
Å	**A**ngstrom (0.1 *nanometers*)
APS	**A**mmonium **p***er***s***ulphate*
AMPR	**Amp**icillin **r***esistance gene*
ATP	**A**denosine **t***ri***p***hosphate*
AU	**A**bsorbance **u***nits*
β	**B**eta
BCA	**B**icin**c**honinic *acid* **a***ssay*
BMC	**B**acterial **m***icro***c***ompartment*
BME	$\beta -$ **M***ercapto***e***thanol*
BCIP/ NBT	5-**B**romo-4-**c**hloro-3-**i**ndolyl **p***hosphate/* **n***itro* **b***lue* **t***etrazolium chloride*
βgal	**B**eta **gal***actosidase*
°C	Degree **C***elsius*
CCM	**C**arbon **c***oncentrating* **m***echanism* (*refers to carboxysome shell proteins*)
CX	(where $X =$ *whole number*) *Location of carbon atom in a chemical compound*
Da	**D**altons
DNA	**D**eoxyribo**n**ucleic **a***cid*
DPN1	**D**iplococcus **pn***eumoniae* **1** (*restriction enzyme*)
ε	Extinction *coefficient*
EDTA	**E**thylene**d**iamine**t**etra**a**cetic *acid*
Etu	**Et**hanolamine *utilisation*
Est5	**Est**erase **5**
FPLC	**F**ast **p***rotein* **l***iquid* **c***hromatography*
g	Force *of* **g***ravity*
GFP	**G**reen **f***lourescent* **p***rotein*
GST	**G**lutathione **S** $-$ **t***ransferase*

Glda	**Gl**ycerol **d***ehydrogenase*
HCL	**H**ydro**ch**loric *acid*
His	**His**tidine
IPTG	**I**so**p**ropyl $\beta - 1 -$ **t***hio***g***alactopyranoside*
IMAC	**I**mobillised **m***etal* **a***ffinity* **c***hromatography*
k	**K**ilo $(x10^3)$
Kd	**D**issociation *constant*
L	**L**itre
LB	**L**ysogeny **b***roth*
M	(prefix) **M***ega* $(x10^6)$
M	(suffix) **M***olar concentration* (mol/dm^3)
m	(prefix) *milli* $(x10^{-3})$
m	(suffix) **m***etre*
MALDI- TOF	**M**atrix **a***ssisted* **l***aser* **d***esorption* **i***onisation* $-$ **T***ime* **o***f* **f***light*
mg	**M**illi**g**rams
mol	**M**ole
MS	**M**ass **s***pectrometry*
MWCO	**M**olecular **w***eight* **c***ut* **o***ff*
n	**N**ano $(x10^{-9})$
NADP+	**N**icotinamide **a***denine* **d***inucleotide* **p***hosphate* $(reduced)$
NADPH	**N**icotinamide **a***denine* **d***inucleotide* **p***hosphate* $(oxidised)$
OD_{600}	**O**ptical **d***ensity at* 600*nm*
OD_{832}	**O**ptical **d***ensity at* 832*nm*
OD_{340}	**O**ptical **d***ensity at* 340*nm*
OD_{280}	**O**ptical **d***ensity at* 280*nm*
ONPG	**o**rtho-**N**itro**p**henyl-$\beta -$ **g***alactoside*
PCR	**P**olymerase **C***hain* **r***eaction*
Pdu	**P**ropane**d**iol **u***tilisation*
Pdc	**P**yruvate **d***e***c***arboxylase*
PAGE	**P**oly**a**crylamide **g***el* **e***lectrophoresis*
PBS	**P**hosphate **b***uffered* **s***aline*
PEG	**P**oly**e**thylene **g***lycol*

pNPB	**p**-**n**itro **p***henyl***b***utyrate*
PDB *or* (*.pdb*)	**P**rotein **D***ata* **B***ank* (*file*)
PTAC	**P**hospho**t**rans**ac**etylase
ppm	**P**arts **p***er* **m***illion*
UV	**U**ltra **v***iolet*
rmsd	**R**oot-**m**ean-**s**quare **d***eviation*
rpm	**R**evolutions **p***er* **m***inute*
Rubisco	**R**ib**u**lose **bis***phosphate* **c***arboxylase* **o***xygenase*
s	**S**econds
SEC	**S**ize **e***xclusion* **c***hromatography*
SDM	**S**ite **d***irected* **m***utagenesis*
SDS	**S**odium **d***odecyl* **s***ulphate*
TAE	**T**ris, **A***cetic acid*, **E***DTA* (*buffer*)
TEMED	**Te**tra**m**ethyl**e**thylene**d**iamine
Tm	**M**elting **t***emperature*
μ	**M**icro (x10^{-6})

Chapter 1

Introduction.

1.1 The History and discovery of Bacterial microcompartments.

1.1.1 The discovery of the carboxysome.

Bacterial microcompartments were first observed in 1956 [1]. Whilst at the time no function was attributed to them, they remained something of a mystery. Subsequently, the presence of polyhedral bodies was sporadically reported in a variety of bacteria [2, 3] (figure: 1.1). In contrast to the vast majority of eukaryotic organelles, these objects featured remarkably regular, polyhedral "membranes", which maintained this regularity following breakages [4]. This indicated that in contrast to the vast majority of eukaryotic cells, which feature organelles enclosed by lipid membranes, this shell was composed of protein.

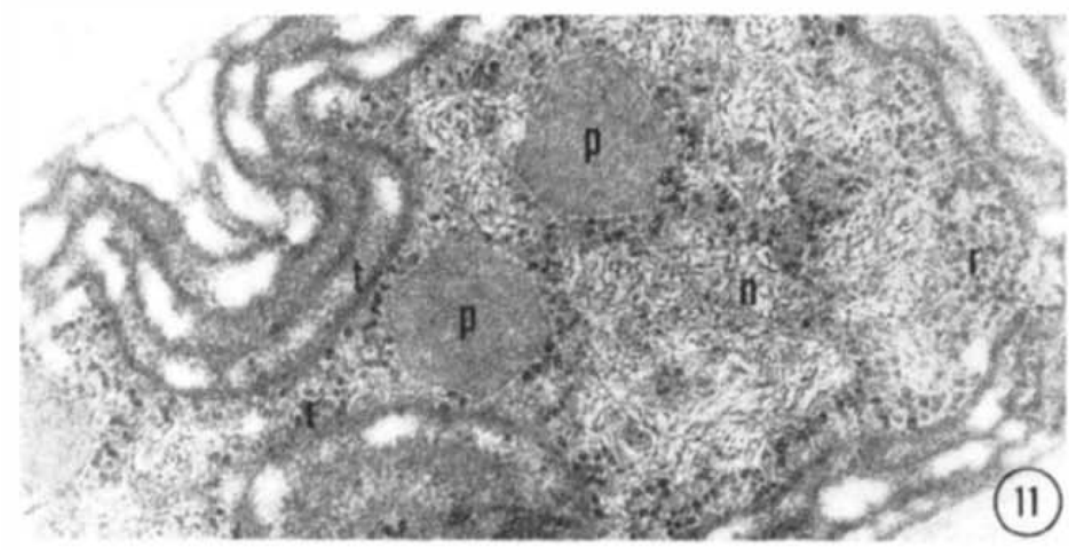

FIGURE 1.1: Electron micrograph of polyhedral organelles of unknown function or purpose. Found within blue-green algae (adapted from Lang (1968) [2]. Many TEM images showing mysterious polyhedral bodies in the bacterial cytoplasm were acquired in the late 1960s-early 1970s. The flat facets and sharp angles visible here are not reminiscent of any eukaryotic organelles.

It was not until the co- isolation of these bodies with the photosynthetic enzyme Rubisco [3], that some involvement in carbon fixation was evident, and the term "carboxysome" was coined. There was speculation as to how the carboxysomes imported CO_2, until the discovery that carbonic anhydrase was also associated with carboxysomes [5, 6]. Carbonic anhydrase converts bicarbonate (which freely permeates the carboxysome shell) into CO_2. CO_2 is then sequestered within the microcompartment lumen, providing a substantially higher concentration within the carboxysome, when compared to the bacterial cytoplasm [7].

The structure of these microcompartments remained elusive for some time, until the elucidation of the crystal structure of a single component of the compartment, CcmK2 [8] (figure: 1.1.1). This structure revealed that CcmK2 was a hexamer (figure: 1.1.1A), which tessellated into flat sheets (figure: 1.1.1C), reminiscent of the flat facets of the microcompartments (which had been previously visualised [2, 3], see figure: 1.1). The characteristics which now typify microcompartment shell proteins were also first identified in this structure. Namely: A polarised pore, at the centre of the hexamer (which permits the influx of the polar bicarbonate, but negates efflux of

CO_2), and the ability of these hexamers to tesselate into flat sheets (which form the facets of the observed polyhedra).

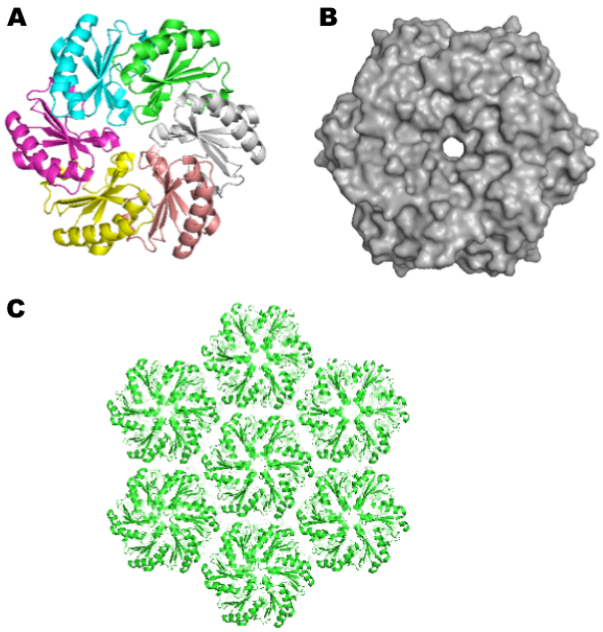

FIGURE 1.2: Structure of the hexameric the shell protein, CcmK2 (PDB: 2A1B). Cartoon representation (A), showing: six identical subunits, each displayed in a different colour. Surface representation (B), showing the six- fold symmetry and clearly visible pore at the centre of the hexamer. The arrangement of hexameric shell proteins into flat sheets (C), as they appear in the crystal lattice, demonstrates how individual shell proteins with six- fold symmetry tessellate into flat sheets.

Mathematically, single hexagons are not capable of forming closed polyhedra, and it was therefore hypothesised that some other molecule must exist to cap the vertices. The missing piece was discovered in the form of CcmL [9] (figure: 1.3). A pentamer, which was a component of the microcompartment shell, alongside the previously identified hexamers. CcmL had eluded detection owing to its unique fold when compared to the previously identified hexamers. Therefore, it could not be identified as a structural component of the BMC through sequence comparison alone. The discovery of these pentameric proteins allowed for the modelling of more complete microcompartment shells where, pentamers cap the vertices of the flat facets

formed by hexamers(figure: 1.3C). The necessity of these pentamers in the formation of complete microcompartments, able to maintain favourable CO_2 concentrations was subsequently verified [10]. Bacteria containing carboxysome mutants, lacking CcmL were not able to successfully sequester CO_2, and showed stunted growth rates when compared to bacteria containing WT BMCs.

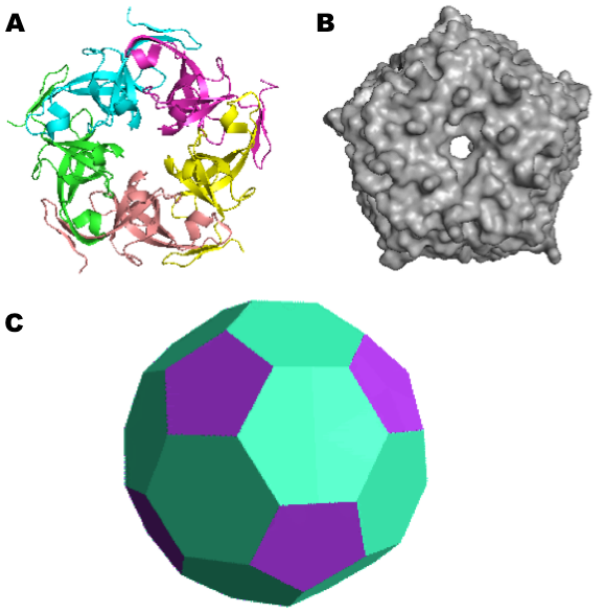

FIGURE 1.3: Structure of the pentameric vertex protein CcmL (PBB: 2RCF). A cartoon representation (A), highlighting the five individual chains is shown. These chains are highly dissimilar to the predominantly α helical hexameric shell proteins, and are primarily formed of β- sheets . A surface representation (B) is also shown, demonstrating the overall shape of the molecule and highlighting the central pore. A simplified representation of a microcomparmtent (C) shows how the pentameric (purple) proteins cap the vertices and the hexameric proteins (green) form the flat facets.

1.1.2 The discovery of the propanediol utilisation (pdu) microcompartment.

Whilst investigations into the structure, functions, and characteristics of carboxysomes progressed; the possibility of other bacteria expressing carboxysome like structures first became apparent, with the discovery of the *pdu* (propanediol utilisation) operon [11, 12, 13] (figure: 1.4). This operon encodes the genes required for the utilisation of 1,2-

PD (1,2-propanediol) as an energy source, and contains enzymes which require an adenosyl- cobalamin (vitamin B_{12}) co-factor. The operon responsible for synthesis of adenosyl- cobalamin, the *cob* operon, and the *pdu* operon were subsequently found to form a single regulon. This regulon is inducible by either 1,2-propanediol or glycerol, and is regulated by the transcription factor *PocR* [12].

Surprisingly, several genes within the *pdu* operon bear remarkable similarity to those which had previously been identified as being components of the carboxysome, in carbon fixing bacteria [13, 14]. The identity of these proteins as being involved in some sort of polyhedral assembly was confirmed, with the observation of carboxysome like particles in *Salmonella typhimirium*. These particles are observable using TEM when grown in 1,2- PD rich media [14]. *S. typhimirium* is incapable of photosynthesis and therefore, despite having highly similar morphologies, these particles could not be carboxysomes. Confirmation that these structures were indeed, non- carboxysomal microcompartments came with the identification of the shell protein PduA. PduA was found to be a key component of these polyhedra [15], and featured the same fold and characteristics as CcmK2.

Whilst catabolic processes are contained within carboxysomes, the presence of genes involved in 1,2- PD degradation, suggested that these compartments were "metabolosomes", enclosing metabolic pathways. Mutants of PduA, featuring a non- polar pore, at the centre of the hexamer, displayed inhibited bacterial growth at high 1,2 PD concentrations [16]. This suggested that, in addition to the substrate concentrating effect taken advantage of by the carboxysome, metabolosomes also act to sequester toxic compounds and intermediates within the microcompartment lumen.

FIGURE 1.4: The Pdu operon. Shell proteins are coloured blue, enzymes involved in propanediol metabolism and co- factor regeneration are coloured green, genes with no known function are coloured black.

1.2 The structural characteristics of shell proteins.

1.2.1 The BMC fold.

BMCs had been routinely observed for 30- 40 years, although the relationship between their constituent proteins and the polyhedral shells remained unclear for some time. The structures of the first two BMC shell proteins; CcmK2 and CcmK4 were solved in 2005 [8] and illuminated how BMCs might form. Here, protein monomers were observed as being wedge shaped, capable of forming hexagonal homohexamers with 6- fold symmetry. Each of these homohexamers featured a central pore (4Å in diameter for CcmK2 and 7Å for CcmK4), lined with polar residues (Lys36 in CcmK2 and Arg38 in CcmK4). The polarity of the pore permits the movement of charged substrates and metabolites across the BMC shell, whilst being impermeable to non- polar molecules such as CO_2 and O_2. The semi- permeable nature of the BMC Shell allows for the maintenance of a concentration gradient, favourable to the actions of the internalised enzymes.

CcmK2 and CcmK4 were the first proteins shown to contain the alpha/beta structural BMC motif, which is now known to be the hallmark of hexameric BMC shell proteins. Whilst amino acid sequence varies to some degree, and there are several permutations of the fold, the 3D structure varies minimally regardless of BMC function. Certain BMC shell proteins feature tandem repeats of the BMC fold and form trimeric pseudo hexamers (PduB, PduT). Others (PduU and EutS) feature circular permutations whereby, despite the amino acid sequence differing, the overall tertiary structure is maintained (figure: 1.5).

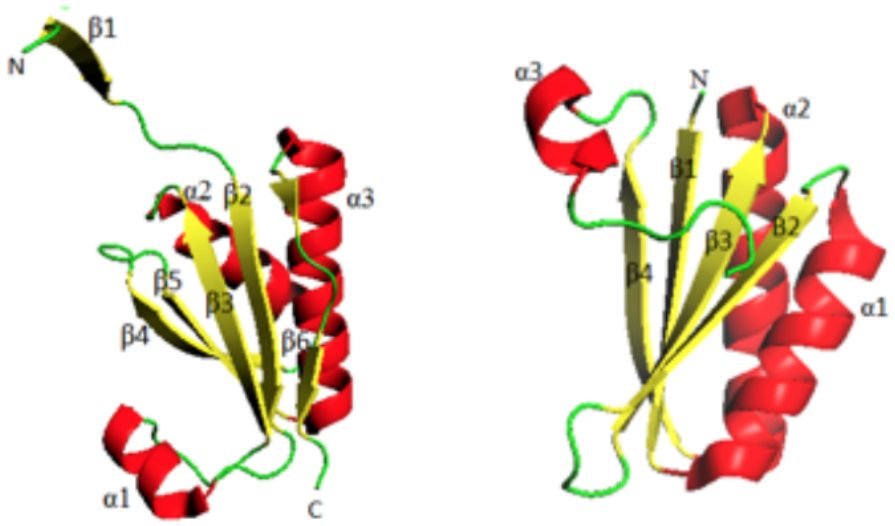

FIGURE 1.5: An overview of the BMC fold. Showing PduU (PDB:3CGI) (A) and PduA (PDB: 3NGK) (B). Whilst the overall wedge shape is maintained and both proteins are capable of forming hexamers, with only marginal differences in overall RMSD, the order of secondary structural elements is completely different.

1.2.2 Shell proteins of the Pdu microcompartment.

The proteins which form the shell of the Pdu microcompartment are highly similar in their structural properties, to those which form the carboxysome. All hexameric and trimeric shell proteins feature the previously described BMC fold, and the pentamers are homologous to CcmL. Despite all sharing a high degree of sequence identity, several shell proteins of the pdu BMC feature unique structural characteristics, which endow them with different functions.

PduU.

PduU was the first shell protein to be successfully crystallised [17] from the Pdu microcompartment (figure: 1.6). The structure revealed that monomers of PduU form hexamers, with each monomer featuring the previously identified BMC fold [8]. PduU shows several key differences and unique properties, when compared to the known structures of carboxysome shell proteins, and indeed the subsequently solved Eut and other Pdu shell proteins. Firstly, hexameric carboxysome shell proteins

show tight interactions between hexamers in the crystal lattice, which is believed to be representative of the biological assembly [8]. In contrast, PduU hexamers do not form these tight layers, but instead form loosely packed arrays, perhaps indicating that PduU associates primarily with other shell proteins. Another key difference between PduU and the structures of other BMC shell proteins, is the extended N-terminus. Here, residues 8- 15 from each monomer interact to form a β- barrel at the central pore of the hexamer. The central pore has often been thought to facilitate movement of substrates and products across the BMC shell. However, in the case of PduU the tight packing of large side chain amino acids within this β- barrel would appear to prevent this. One possibility is that the unusually large concave surface facilitates the docking of interior components of the BMC to the outer shell. It is also permissible that under certain circumstances, the large pore lining residues could move outwards allowing for movement of molecules across the shell surface.

Interestingly, bacteria deficient in PduU are still capable of forming empty non-aberrant BMCs [18], and PduU knockout mutants still form phenotypically similar BMCs [19]. This suggests that PduU may play a role in transport of small molecules, or in facilitating interactions with encapsulated enzymes, rather than an essential structural role.

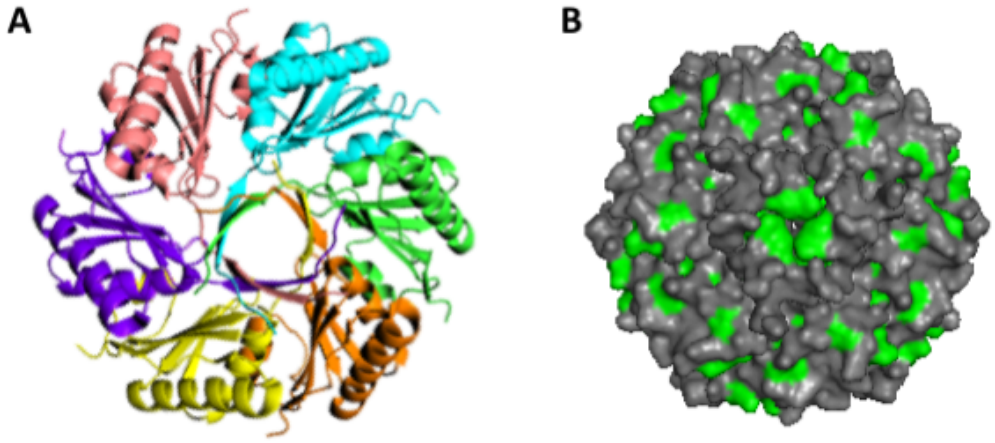

FIGURE 1.6: Cartoon representation of the crystal structure of PduU (PDB:3CGI). Six individual, identical chains, showing six- fold symmetry are highlighted in different colours (A). The β- barrel located at the central pore is also shown, with the C- terminus of each subunit contributing one of the six β- strands required. In addition, a surface representation with hydrophobics coloured in green is also shown (B), demonstrating that the tight packing of hydrophobic residues in the pore would prevent the movement of small molecules across the PduU hexamer.

PduT.

Whilst most BMC shell proteins are hexamers, some form pseudo- hexameric trimers. Each monomer within these trimers features two BMC folds, which contribute to the final hexagonal shape of the protein, maintaining the six- fold symmetry seen in hexamers. PduT is one such pseudo- hexamer [20] (figure: 1.7), and features an iron Sulphur cluster at its central pore. Sulphur's are contributed from Cys38 from each monomer, and these cysteines can each bind one of four irons within the iron Sulphur cluster. Therefore, one iron remains unbound, potentially allowing for an interaction with another protein, providing a mechanism for interactions between shell proteins and encapsulated enzymes.

The PduT trimer is also relatively flat when compared to other BMC shell proteins, providing a central pore that is easily accessed from either side. This feature makes it an ideal candidate for single electron transfer across the BMC shell. As with PduU

a loss of PduT does not prevent the formation of empty BMCs [18], or notably effect the overall BMC structure in PduT knockout mutants [19].

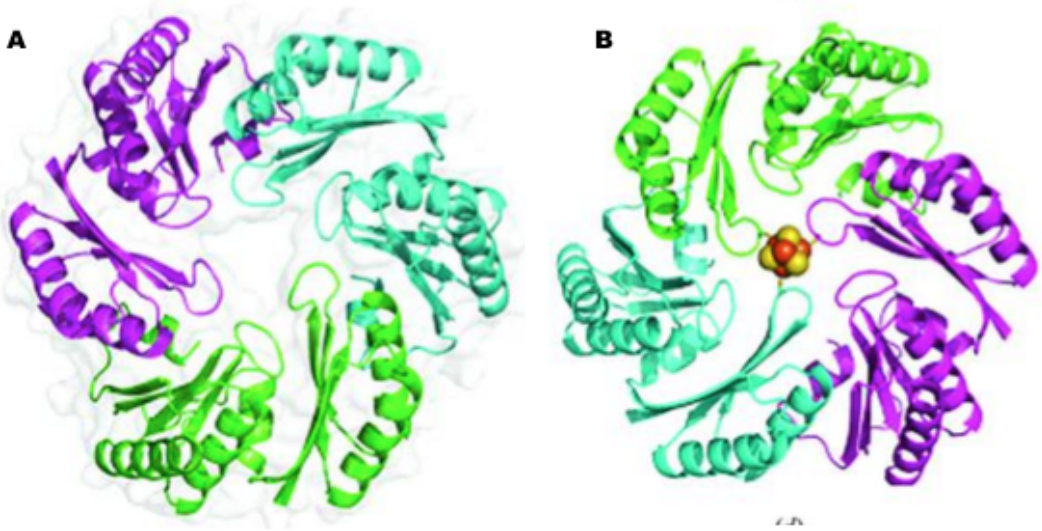

FIGURE 1.7: Cartoon representation of the crystal structure of PduT (PDB: 3PAC)(adapted from Pang (2011) [20]. Three individual, identical chains are highlighted in different colours (A), each composed of two near identical BMC domains, preserving the six- fold symmetry. A superimposed 4Fe- 4S cluster is shown (B), demonstrating where this cluster is likely found, at the central pore of the trimer.

PduB.

Another Pdu shell protein featuring a tandem BMC repeat, and showing pseudo-hexameric symmetry is PduB (figure: 1.8). PduB is unique in that it is the only BMC shell protein which has been shown to interact with a native substrate [21]. The crystal structure of PduB shows its ability to bind glycerol molecules at the interface between monomers within the trimer, and at the central pore of the pseudo hexamer. The glycerols bound at the monomer interface are suggested to play a role in promoting or inhibiting access to the central pore, by locking the central loops which line the pore. This would make PduB the only known ligand gated channel within BMCs at present. An additional variant of PduB, dubbed PduB' results from there being two translation start sites on the polycistronic message [22], resulting in this variant lacking the 37 N- terminal residues of PduB. The significance of these residues is largely unclear, although BMCs exclusively expressing PduB' are

deficient in the enzymes usually localised to the lumen [23]. This would therefore indicate some roles in cargo encapsulation.

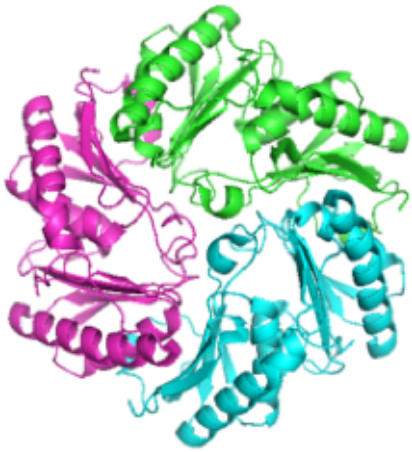

FIGURE 1.8: Cartoon representation of the crystal structure of PduB (PDB: 4FAY). Three individual, identical chains are highlighted in different colours, each composed of two BMC domains, preserving the six- fold symmetry are shown. The pore of the PduB trimer is lined with polar residues.

PduA.

PduA is a hexamer [24] (figure: 1.9) and one of the most abundant shell proteins within the Pdu BMC [18]. It had previously been shown to interact with the majority of other shell proteins [18], and so is purported to act as a scaffold for assembly of the Pdu BMC. The PduA hexamer is concave with a central pore of 8 Å. The polarity of this pore is of particular importance in regards to the import of substrate molecules and sequestration of toxic intermediates. Where the serine usually located at position 40 in the pore is mutated a to non- polar residue, bacterial growth is severely inhibited [16]; Presumably as a result of the efflux of both substrate and toxic molecules into the bacterial cytoplasm. The presence of PduA is essential in attaining higher order structures. This is evidenced by PduA deletion mutants, which form aberrant filamentous structures within the bacterial cytoplasm [19]. Remarkably, when PduA is expressed alone it is capable of forming highly ordered, hollow nanotubes with a diameter of 20.4nm [25], which can be seen under TEM within the bacteria.

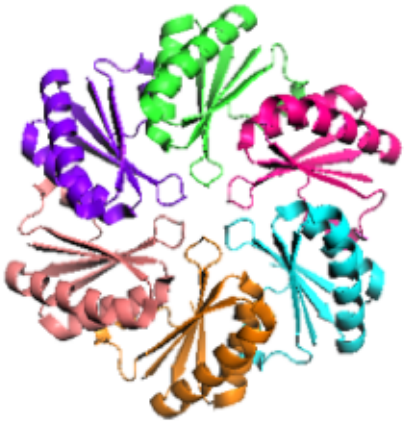

FIGURE 1.9: Cartoon representation of the crystal structure of PduA (PDB: 3NGK). Six individual, identical chains are highlighted in different colours. Each chain consists of a single BMC domain and the hexamer features six- fold symmetry. The pore lining residues of PduA are polar, allowing for influx and efflux of polar molecules.

PduN.

There is limited experimental data available on PduN, however sequence homology with known pentameric proteins CcmL, OrfA and EutN suggests that it is a pentameric shell protein. Such proteins are necessary in addition to hexamers in icosahedral structures to cap the vertices of the microcompartment and prevent any "leakiness" [10].

PduJ.

PduJ is near identical to PduA in terms of sequence homology (90%), and RMSD of the crystal structures [26]. Despite this, whilst PduA is critical in the transport of 1,2- PD across the BMC shell, PduJ has no apparent role in transport. Even more surprising is that where the gene positions of PduA and PduJ are swapped, PduJ acquires the 1,2- PD transport functionality of PduA. Indicating that the ordering of shell proteins within the *pdu* operon, has a direct influence on their role in the BMC. Additionally, in ΔPduA mutant BMCs (where PduA is absent), the relative amount of PduJ to other shell proteins increases dramatically [27].

PduK.

PduK is perhaps the most interesting Pdu shell protein with no known structure. At 150 residues in length it is comparable in size to the trimeric, pseudo- hexameric shell proteins, yet has only one BMC motif. The function of this large C- terminal extension is unknown, although it is weakly similar to the EutK C- terminus (15%). The EutK C- terminus adopts a helix- turn- helix conformation, and is homologous to proteins involved in nucleotide binding [28]. EutK is currently unique amongst BMC shell proteins, in that it exists as a monomer in solution, and is hypothesised to form heterohexamers with other shell proteins. Preliminary data however, suggests that PduK exists as a hexamer in solution [29].

The presence of four proximal cysteines, at the C- terminus of PduK (which are not present in EutK) has led to speculation regarding the presence of an iron- Sulphur cluster, similar to PduT. Although, when purified aerobically no brown colour, or absorbance peak at 420nm are seen, as is typical for Fe-S proteins. Anaerobic purification would be required to eliminate the possibility of an oxygen labile Fe-S cluster.

1.3 Higher order structures formed by shell proteins.

For some time, how constituent shell proteins assembled into microcompartments was largely unknown. Crystallographic data was able to elucidate the homo- oligomeric shell protein interactions within the flat sheets, often seen in crystal structures. Although, how different shell proteins interact with one another, and how those interactions influence the size and shape of microcompartments remained unclear.

In addition to the large, hetero- oligomeric microcompartments formed by the shell proteins, several shell proteins form large homo- oligomeric complexes; in the form of tubes [25] or sheets [17]. PduA has been observed to tesselate into flat sheets, or curved nano- tubes, a process facilitated by high protein concentrations and low NaCl concentrations. The residues responsible for the hexamer- hexamer interaction (Lysine 26, Valine 51 and arginine 79) in PduA are, unsurprisingly located at the hexamer edge, and so are able to bridge the gap between the adjoining hexamers

(figure: 1.11). The primary residue responsible for tessellation is a lysine at position 26 in PduA, which is conserved across the vast majority of known shell proteins. These lysines form an antiparallel interaction, where the amino group of one lysine forms a hydrogen bond with the backbone carbonyl of the other. Previous work mutating this lysine, demonstrates a reduced propensity for tessellation [25] where it is substituted for alanine. When the lysine is substituted for aspartate the ability to tesselate is abolished completely. Within the *Citrobacter freundii* pdu BMC, this lysine is conserved across all BMC domains (figure: 1.10), potentially allowing for interactions between different shell proteins in the formation of the BMC facets.

FIGURE 1.10: Alignment of all BMC domain containing proteins from the Pdu microcompartment. Lys 26 is conserved across all proteins and domains within those proteins. Val 51 is largely conserved, or a similar hydrophobic residue such as leucine is present. Arg 79 is conserved across all shell proteins, with the exception of PduU and PduB. Both BMC domains of pseudo- hexameric shell proteins are included, as they are not identical.

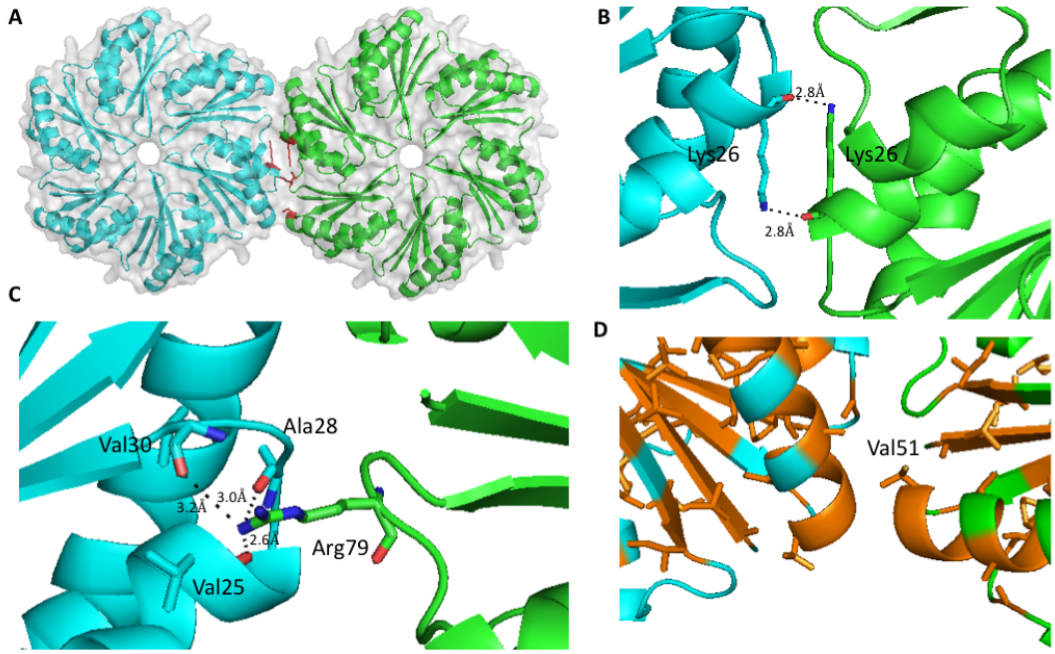

FIGURE 1.11: A pair of tessellating PduA hexamers (A), each hexamer is shown in either green or cyan with interfacial residues involved in tessellation coloured in red. Several zoomed views of the hexamer- hexamer interface are shown, demonstrating the key roles of Lys 26 (B), Arg 79 (C) and Val 51 (D) in hexamer tessellation. Both Lys 26 and Arg 79 participate in electrostatic interactions with residues in the adjacent hexamer. Val 51 is involved in hydrophobic packing with several hydrophobic residues in the adjacent hexamer.

Shell protein nanotubes.

Nanotubes are known to be formed by both PduA and PduB [25, 30]. PduA nanotubes are frequently observed at around 20.4nm diameter whereas, PduB nanotubes are substantially larger at 63nm. PduA nanotubes are hypothesised to form where the flat plane between two PduA hexamers is distorted by 36°(figure: 1.12). This results in the cleft present at the interface of two PduA hexamers contracting. This model of tube formation is supported by the fact that gold nanoparticles (which bind to the His- tag of recombinant proteins) will bind to the exterior of a PduA nanotube [30]. Therefore, the His- tag and N- termini of PduA are on the exposed, or outer surface of the nanotube. The existence of these nanotubes is curious, as they are only found where single shell proteins are expressed in isolation; suggesting

that there is some preference between shell proteins during assembly, which facilitates the formation of compartments with flat facets, rather than the curvature seen in nanotubes. A lack of curvature in the crystal structures of single shell proteins is also apparent, and it is unclear why these proteins tend towards a planar assembly in the crystal lattice, yet overwhelmingly tend towards curved assemblies when assessed by TEM in vivo and in vitro.

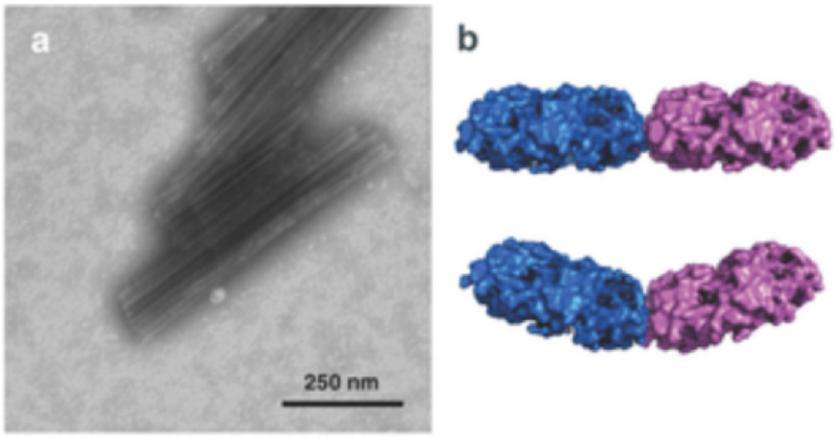

FIGURE 1.12: Electron micrograph of the tubes formed by PduA (A) and a computationally produced model demonstrating how distortion of the planar interaction between two PduA hexamers might contribute to the formation of such a complex. Adapted from Uddin et. al (2018) [30]

Heterologous interactions between shell proteins.

Whilst residues important for tessellation are largely conserved across shell proteins in the Pdu BMC, there does appear to be some discrimination in terms of which shell proteins preferentially interact with each other (table: 1.1). Pull down studies have demonstrated a particular propensity for PduA to interact with the majority of other shell proteins [18]. This information, combined with PduA being one of the more abundant shell proteins in the *C. freundii* Pdu BMC, at around 20%; indicates that PduA may be a hub-like protein to which other shell proteins adhere. Surprisingly, the trimeric shell protein PduB has been reported as being more abundant than PduA (30%) [18], and yet has only been reported to interact with itself and PduA. Interestingly, within the *S. typhimirium* BMC PduA and J are the major

TABLE 1.1: Summary of the known interactions between different shell proteins, as determined using pull down assays [18]. PduA notably interacts with most other shell proteins.

	prey								
Bait Protein	PduA	PduB	PduJ	PduK	PduN	PduU	PduT	PduV	PduM
PduA	NA	Y	Y	Y	Y	Y	N	N	N
PduB	Y	NA	N	N	N	N	N	N	N
PduJ	Y	N	NA	N	N	N	N	N	N
PduA	Y	N	N	NA	N	N	Y	N	Y

shell proteins [27], with PduB being present in a lower amount as compared with *C. freundii*. Regardless of the species from which the BMC is obtained it is apparent that PduA is a major structural component of the Pdu BMC. The small amounts of other shell proteins may serve more niche roles, such as import/ export of various molecules, and/ or binding to specific encapsulated enzymes.

The formation of whole microcompartments.

There has been much work on the structures and function of individual shell proteins. Although, there is little information available on the way these proteins assemble into whole microcompartments. No structures have been solved containing multiple Pdu shell proteins, and so the way that these hetero- oligomers form can only be speculated or modelled. In spite of this, parallels can be drawn between the Pdu BMC, and the recently solved structure of a comparatively small (6.5 MDa) BMC. This BMC is produced by *Haliangium ochraceum*, and has no known function [31]. The structure contains shell proteins homologous to PduA, T and N (figure: 1.13). The hexameric protein Hoch 5815 is homologous to PduA. The three very similar Hoch 5816, 5812 and 3341 proteins are similar to PduT in their overall fold, but differ in their lacking the cysteines required to form an Fe-S cluster. The pentamer Hoch 5814 is homologous to PduN . The way these shell proteins assemble into whole microcompartments is interesting, with shell proteins having clear preferences for which other shell proteins they will interact with. Despite the trimers (Hoch 5816, 5812 and 3341) forming flat sheets in the crystal lattice, with contacts to other like proteins, no such interactions occur in the structure of the whole microcompartment. The trimers interact exclusively with hexamers, and these hexamers

are able to interact with trimers, pentamers or other hexamers. This data is consistent with previous observations that PduA is critical for microcompartment formation, and forms a major component of the Pdu BMC. This does not however, elucidate the role of PduB, which is abundant in the Pdu microcompartment. PduB is known to be critical for Pdu microcompartment formation [18], but has no homologue in the Hoch microcompartment. The angles formed between certain shell proteins in the Hoch microcompartment are also interesting. The only proteins capable of forming flat interfaces, where the angle between adjacent proteins is 180° are the hexamer-hexamer interfaces. This is reminiscent of the flat facets of PduA within the Pdu microcompartment. The hexamer- hexamer interface is quite dynamic and also forms at 150° angles, accounting for the discrepancies seen between PduA nanotubes and PduA sheets. Indicating that, the protein is capable of forming, and is stable in either configuration (180° or 150°). The hexamer- trimer and hexamer- pentamer interfaces are present in only a single conformation, and form at angles of 155° and 150° respectively. These plane distortions are responsible for the curvature in the shell, and allow for the sealing of the microcompartment. Figure: 1.14 shows the curvature formed and places the C- terminus (convex side) of the PduA like protein to be on the exterior of the microcompartment. A result consistent with previous observations on PduA nanotubes [30].

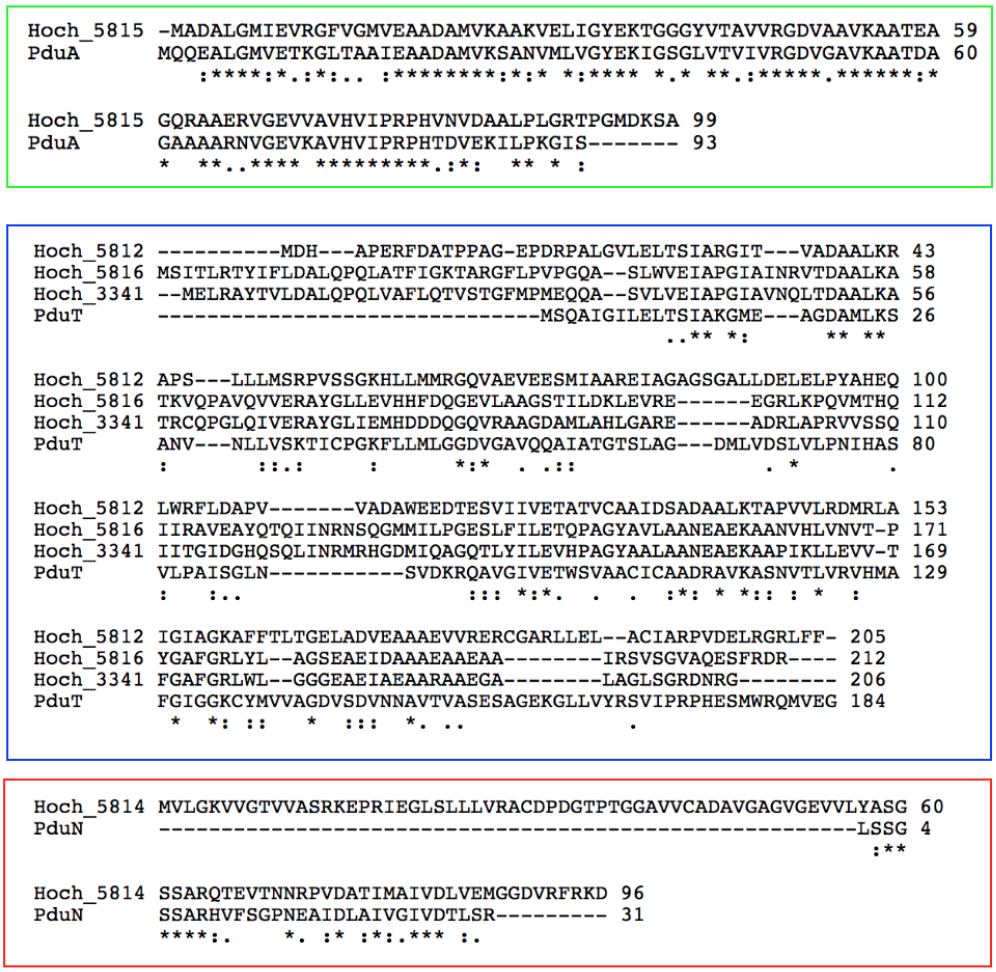

FIGURE 1.13: Alignment of Pdu shell proteins and those found in the *H.Ochraceum* BMC. The hexamers are homologous to PduA (highlighted in green). The trimers are homolgous to PduT (alignment highlighted in blue). The pentamers are homologous to PduN (highlighted in red). Within the *H.ochraceum* BMC there are no homologues to PduB, or PduU.

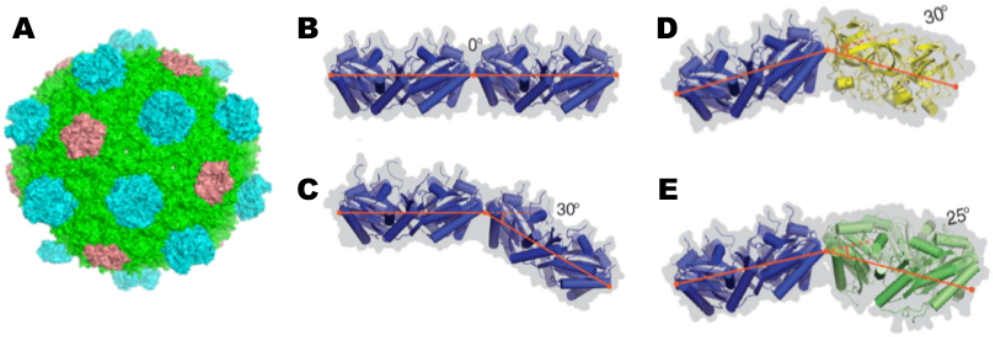

FIGURE 1.14: Surface representation of the whole microcompartment crystallised from H. Ochraceum (A). Additionally the interactions and angles produced between hexamers and hexamers (C,D) hexamers and pentamers (D) and hexamers and trimers are shown (E). (A) was generated using pymol from the deposited co -ordinates PDB: 5V74. (C), (D) and (E) were adapted from Sutter et. al (2017) [31]

1.4 Overview of the Pdu BMC metabolic pathway.

Since the discovery of the Pdu microcompartment, much work has been done on the isolation and characterisation of the key enzymes in the pathway, which are present within the microcompartment. In brief, the Pdu pathway operates to produce propionate, which can then be used as a carbon source in a variety of bacteria [32]. This pathway creates propionaldehyde as an intermediate, a molecule with a variety of highly cytotoxic properties [33]. It is therefore necessary to encapsulate the pathway within the microcompartment, in order to protect the bacteria from the cytotoxic effects. There is some speculation as to whether the entire pathway is encapsulated, or just specific enzymes which require direct access to propionaldehyde.

1,2- PD is initially converted into propionaldehyde by the PduCDE complex, A cobalamin dependant diol dehydratase [34, 35]. The complex is a hetero- trimeric dimer ($(CDE)_2$), and is closely related to the glycerol dehydratases found in the cytoplasm of many bacteria. Adenosyl- cobalamin is used as a proton acceptor in this reaction, to produce the products: propionaldehyde, cobal(II)amin and an adenosine radical [36]. The regeneration of the adenosyl- cobalamin cofactor is required

for PduCDE to continually produce propionaldehyde. Cofactor regeneration is undertaken by the enzymes: PduS [37] and PduO [38], as well as the PduGH complex in an ATP dependent reaction [39, 40].

Propionaldehyde produced by PduCDE may be converted into 1- propanol by PduQ [41] and exported, terminating the pathway. Alternatively, it may continue down the pathway to be converted into propionaldehyde by PduP [42]. The actions of PduP and PduQ are cyclical whereby, the reaction catalysed by PduP reduces NAD+ into NADH, which can then be oxidised by the reaction catalysed by PduQ. This mechanism of cofactor regeneration is particularly useful for large, semi- permeable structures such as BMCs, as it negates the requirement for large molecules to be able to pass the shell, thereby maintaining the integrity of the compartment.

Propionyl- CoA produced by PduP is converted into propionyl- phosphate by the phosphotransacetylase (PTAC), PduL [43, 44]. PduL recovers active CoA in the process, which can then be used by PduP in the synthesis of more propionyl- CoA, again demonstrating the linked nature of the enzymes in this pathway. Propionyl-CoA is then converted into propionate by the propionate kinase, PduW [45]. A process which transfers a phosphate to the ADP produced by PduGH, and thereby eliminates the necessity for import of ATP from the cytoplasm. While there is no evidence for the encapsulation or association of the enzyme PduW with the microcompartment, it seems a missing piece of the puzzle as to how microcompartments import the large cofactors required by the constituent enzymes. The import of large molecules into the BMC lumen has been a point of interest for some time. A commonly cited mechanism involves the pore lining residues of certain trimeric shell proteins opening to facilitate the passage of large molecules [46]. However, in order for the BMC to maintain its selective permeability, two of these proteins must be stacked. Here, both are never in the open conformation at the same time and so function as a kind of airlock [47]. Whilst this mechanism may be present in some BMCs, there are no Pdu shell proteins which are known to both stack in pairs and which feature pore lining residues capable of moving to allow passage of large molecules. It is perhaps more likely that enzymes which are currently not known to be internalised, are in fact associated to some degree and large molecules are recycled by

these enzymes. Figure: 1.9 shows that every large substrate within the Pdu pathway is recycled by some other element of the pathway, and the only molecules which require entry to the BMC (should all enzymes in the pathway be internalised), are the small, polar molecules: Phosphate and 1,2- PD.

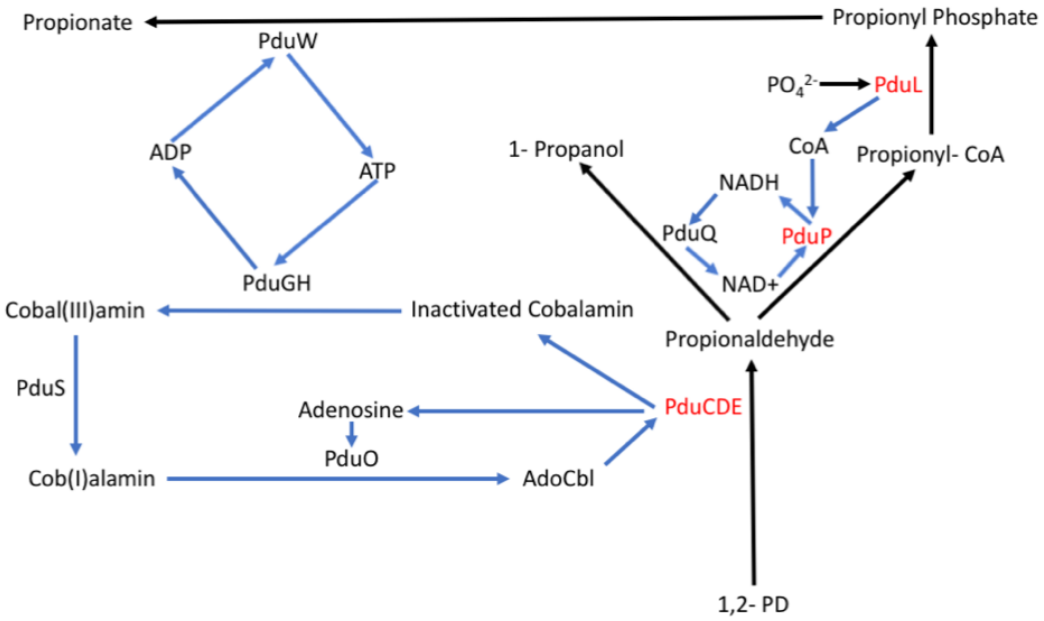

FIGURE 1.15: A diagramatic representation of the metabolic processes in the 1,2-propanediol utilisation microcompartment. Enzymes with known encapsulation sequences are shown in red and pathways by which large molecules are recycled within the compartment are shown in blue.

1.5 Encapsulation of enzymes within the bacterial microcompartment.

The encapsulation of internalised enzymes within the Pdu microcompartment is accomplished through the use of 20 residue signalling sequences. These sequences are present on the N- terminus of encapsulated proteins, and are known to be present on at least PduP [48], D [49], L [50], and S [37].

The role of the 18 N- terminal residues of PduP (P18) in targeting the enzymes to the

BMC was the first interaction to be established [48]. Here, removing the 18 N- terminal residues was shown to eliminate the ability of PduP to associate with BMCs. Fusion of P18 onto proteins not usually associated with BMCs (GFP, GST and maltose binding protein), resulted in these proteins being co- isolated with purified BMCs. Association of PduP with the BMC also permits the encapsulation of PduQ [41], which is associated with PduP through its N- terminus. More recent studies using pull- down assays have shown that PduP interacts specifically with PduA, PduJ and PduK [51].

In addition to the role of P18 in enzyme encapsulation, the N- terminus of PduD (D18) has also been implicated [49]. Where the *PduC*, *PduD* and *PduE* genes are deleted from *E. coli* and replaced individually, only PduD is capable of associating with the BMC when expressed alone. In addition, where either PduC or PduE are expressed alongside PduD, they are also found to be encapsulated. This is not true when expressing either PduC or PduE in the absence of PduD; suggesting it is PduD responsible for encapsulation of the complex. Deletion of the 35 N- terminal residues of PduD has detrimental effects on the ability of the PduCDE complex to associate with the BMC. Therefore, as with PduP, it is the N- terminus which plays a role in targeting. As with PduP, fusion of the N- terminal residues of PduD onto the N - terminus of GFP or GST results in their co- localisation within the BMC.

Evidence suggests PduL is a luminal enzyme associated with shell proteins through its twenty N- terminal residues [50]. Deleting even the first five of these residues severely effects the ability of PduL to co- associate with purified BMCs, and fusing these twenty N- terminal residues to GFP results in the encapsulation of GFP within the BMC. Intriguingly, the PduL signalling peptide (L20) is the only BMC signalling peptide not predicted to form an amphipathic α- helix [52].

One of the less indiscriminate interactions documented between encapsulated enzymes and shell proteins, is the interaction between PduT and PduS, as PduS does not appear to interact with any other shell proteins [37]. When PduT and PduS are co- purified with an N- terminal his- tag on PduT, both proteins are eluted together. However, where this His- tag is present on the N- terminus of PduS, no such interaction occurs. This is likely a result of the his- tag interfering with the residues

involved in binding, which are present in the PduS N- terminus. Therefore, as with other encapsulated enzymes, it seems it is the N- terminus of PduS which facilitates its encapsulation.

PduV (a GTPase suspected of existing on the exterior of the microcompartment) has been shown to interact with PduU experimentally, and this interaction has been modelled [52]. Unlike other interactions mentioned, this does not appear to result from the presence of an extension peptide, but instead is postulated to form from an interaction between the PduU β- barrel and residues 15- 35 of PduV. Forming a 1:1 ratio of PduU hexamer and PduV monomer. Whether this binding event is due to a unique relationship between the PduU β- barrel and a tertiary structural element of PduV is unknown. Fusion of the 35 N- terminal residues of PduV onto a heterologous protein such as GFP could help establish whether this is the case.

1.5.1 Encapsulation of enzymes within the core of a Bacterial microcompartment are integral in its formation.

In carboxysomes, it has been demonstrated that not only do the signalling peptides play a role in binding the enzymes to the shell proteins, but are in fact integral to the formation of the entire BMC [53]. Here, a signalling peptide present on the small subunit of CcmM (an encapsulated carbonic anhydrase), was shown to be essential for Carboxysome biogenesis, and removal of this sequence abrogated Carboxysome formation. Where this sequence was present, it could be seen to co- precipitate with RuBisCO into a "Pro- Carboxysome" and the microcompartment shell then formed around this. This study has demonstrated that how extension peptides bind to shell proteins, and how the microcompartment forms are two inextricably linked processes.

1.6 Paradoxical aspects of enzyme encapsulation in the Pdu microcompartment.

There is evidence that the 14 C- terminal residues of PduA are involved in binding P18 [51], as a truncated PduA mutant lacking these residues does not appear to interact with PduP. Scanning alanine mutagenesis has been employed to identify specific residues in the N- terminal region of PduP, and the C- terminal region of PduA which are responsible for this interaction. Glu7, Ile10 and Leu14 within PduP and His81, Val84, and Leu88 within PduA were demonstrated to be highly important for PduP encapsulation, and interaction with PduA. A recently proposed model [52] of the PduA- PduP complex purported that, a conserved arginine residue (at the centre of all known Pdu BMC signalling peptides) formed a salt bridge with a conserved glutamate residue, within the BMC domain of PduA. Hydrophobic interactions are also speculated to play a role in this interaction. There are however key issues with this model, as well as the idea that the C- terminus of PduA interacts directly with the N- terminal signalling peptide. Crucially, key evidence has recently arisen [31, 30] (figures: 1.12 and 1.14) that the C- termini of PduA like proteins are present on the exterior of the BMC. This observation is directly antithetical to the proposed models, as the primary goal of the Pdu BMC is to sequester various molecules produced by the internalised enzymes. Something which would not be feasible, should the enzymes be presented on the microcompartment exterior, as this model would suggest. Another key issue is the purported conservation of a central arginine in the centre of signalling sequences. Whilst this residue is indeed conserved in *S. typhimirium*, it is not in *C. freundii*, as well as several other species.

In light of this evidence, it seems prudent that alternate mechanisms for enzyme encapsulation should be explored. In order to better and more precisely manipulate BMCs for synthetic applications, a clearer understanding of cargo encapsulation is required.

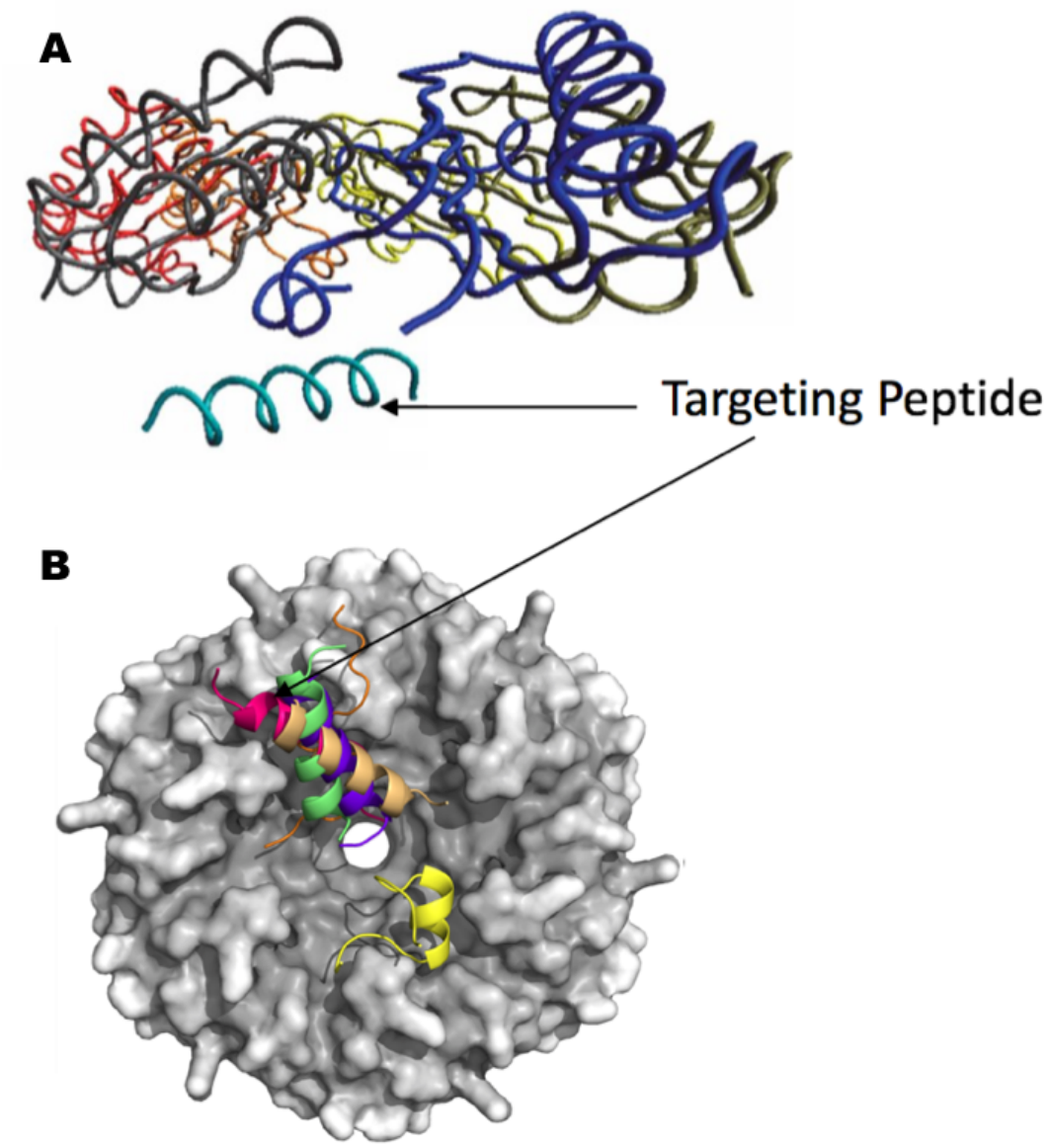

FIGURE 1.16: The previously proposed mechanisms of signalling peptides binding to shell proteins (adapted from Fan et. al. (2012) (A) [51] and Jorda et. al. (2015) (B) [52]). Both of these models show the critical importance of the C- terminus in the interaction, although evidence now suggests the C- terminus is positioned on the exterior of the microcompartment.

1.7 Uses and applications of microcompartments.

1.7.1 Biotechnology

The existence of a proteinaceous shell, which is selectively permeable to a variety of molecules may prove an invaluable tool in biotechnology. The fact that signalling sequences, which allow for the targeting of non- native enzymes to the microcompartment lumen exist, is an additional boon. It is speculated by tethering various components of an enzymatic pathway to shell proteins, the overall rate of product synthesis will be enhanced. This effect would be produced by a pseudo- substrate channeling mechanism, and by reducing the diffusion distance of metabolites between their respective enzymes.

Several heterologous enzymes have been successfully directed to Pdu micro compartments. These include: aldehyde dehydrogenase (Adh) and pyruvate decarboxylase (Pdc); the two enzymes necessary for conversion of pyruvate to ethanol. Where Adh and Pdc are targeted to the BMC lumen via P18 and D18 signalling peptides respectively, a greater yield of ethanol is produced when compared to Adh and Pdh in solution [54].

More recently, β- galactosidase (βgal), Esterase (Est5) and cofactor-dependent glycerol dehydrogenase (Glda) were directed to the Pdu BMC lumen via the P18 signalling peptide [55]. Several interesting observations were made during this study. Firstly there appears to be a currently unknown factor which effects how readily enzymes are packaged into the BMC lumen. Here, despite all enzymes featuring the same signalling sequence Est5 accumulated in 63 fold greater amounts when encapsulated, compared to in solution. However, both βgal and Glda were present at 23 fold greater amounts. The reason for this disparity remains unclear and may be key to understanding exactly how BMCs encapsulate their targets.

The effect of the N- terminal signalling peptides on catalytic activity also appears to be rather selective. Fusion of P18 onto βgal enhanced specific activity to 290%, however, the specific activity of Est5 was reduced to 49%. This shows the effect of the extension peptide varies dependent on the enzyme to which it is tethered, and likely

has different effects dependent on tertiary structure, oligomeric state, translation, transcription etc. Trial and error may be the only viable methodology to ascertain whether an enzyme/ enzymatic pathway would benefit from being encapsulated in a BMC.

Encapsulation also appears to shield enzymes from environmental, but not physical stress; as BMC associated βgal maintains 87% of original activity at pH4. Whereas, the activity of free Bgal drops to 13%. This protection however, does not extend to temperature, as both BMC associated and free βgal are rendered inactive at temperatures of 58°C. An effect likely a result of both BMC and Bgal denaturation.

Comparisons between the substrate availability for both Bgal and Est5, dependant on whether they are encapsulated is interesting. The substrate for βgal (ortho-Nitrophenyl-β-galactoside(OPNG)) is is more readily available where βgal is encapsulated. Whereas, the availability of the substrate for Est5 (p-nitro phenylbutyrate(pNPB)) is reduced. pNPB is substantially more hydrophobic than OPNG, suggesting the BMC shell acts as a barrier to hydrophobic molecules. A theory supported by the observation that PduA Ser40 > Ala/Leu mutants reduce polarity of the central pore, and enable diffusion of the less polar Propionaldehyde, whilst reducing the influx of the more polar 1,2- Propanediol [16].

1.7.2 Medicine

Bacterial microcompartments containing the pathway for ethanolamine metabolism have been isolated from *E. Coli* in patients with urinary tract infections [56]. Metabolism of ethanolamine, like propanediol produces a toxic aldehyde intermediate: Ethanaldehyde. Sequestration of the enzymes involved within a microcompartment shields the bacteria from the toxic effects of the aldehyde. There are comparatively few nutrients present in urine, and so the ability to use the rather ubiquitous ethanolamine as a carbon and nitrogen source, may confer a substantial advantage to bacteria thriving in the urinary tract. This is demonstrable where bacteria containing the *Etu* operon produce BMCs when grown in artificial urine, conferring a growth advantage.

The presence of microcompartments in pathenogenic bacteria is of particular interest. Knowledge of how these compartments assemble, as well as the key interactions which promote and maintain assembly, will be crucial for potential antimicrobial agents which target microcompartments. Drugs able to prevent BMC assembly would eliminate the advantages conferred to bacteria which contain microcompartments.

There has been some speculation that microcompartments may make suitable drug delivery systems [57, 58]. Enough interest has been generated that several patents pertaining to the use of microcompartments as drug delivery systems have been filed [59, 60]. The large microcompartment lumen would be suitable for a variety of large molecule or indeed small molecule therapies. The selectively permeable nature of BMC shells is also of particular interest, especially in regard to the slow, long term release of therapeutic molecules. The ability to alter the polarity of the central pore in shell proteins would also be of particular benefit in this application; Allowing for the control of the efflux of drugs into the environment. Two further key aspects of using BMCs as drug delivery systems which remain unclear are:

1. The potential for the BMC to produce an immune response in patients, thereby rendering them unsuitable as drug delivery vehicles.

2. Whether or not a controlled disassembly of BMC's is possible in order to release the desired molecule.

In regard to the first issue, whether or not BMC shell proteins elicit an immune response in humans is unclear. Although work on PEGylation [61] with other therapeutic molecules and drug delivery systems has shown that: conjugating PEG to the molecule of interest reduces unwanted interactions by disrupting surface electrostatics and thereby reducing the immunogenicity of the molecule. This technique may have future roles to play within the harnessing of using BMC's as drug delivery systems.

The second issue is one with no clear answer. Whilst BMC's are by no means permanent fixtures of the bacterial cytoplasm, they are quite stable structures, capable of

persisting for several weeks. Therefore, controlled disassembly to release the therapeutic molecules would be required for use in drug delivery. The development of molecules which are able to disrupt the interfaces between shell proteins would therefore be required, and so a greater knowledge of the inter- shell protein interactions is essential.

1.8 Goals of this thesis.

Given the inconsistencies between recent experimental observations regarding the topography of shell proteins within microcompartments, and current thoughts on how cargo encapsulation is attained; It would be of great benefit to explore the interactions between shell proteins and cargo enzymes. The aims of this thesis are therefore:

1) Investigate the tessellation of PduA in solution to determine whether an interface is present which might serve as a binding site for encapsulated enzymes.

2) Identify whether the phosphotransacetylase, PduL interacts with the shell proteins PduA and PduK. Using techniques including Native PAGE, fluorescence quenching and Circular dichroism.

3) Investigate the binding constants and stoichiometry of the enzyme- shell protein complexes to better understand how the complex forms and to aid the generation of computationally derived models.

4) Produce a model of the shell- protein- enzyme complex of PduA and PduL. Either through X- ray crystallography or using in silico techniques such as HADOCK [96] or FLEXPEPDOCK [84]

These aims are explored in the following chapters, using a variety of techniques, including: Native PAGE, mass spectrometry, size exclusion chromatography, X- ray crystallography, electron microscopy, circular dichroism, fluorescence quenching and computational modelling.

Chapter 2

Materials and Methods.

2.1 Materials.

TABLE 2.1: Materials and suppliers used throughout this thesis.

Material	Supplier
kb DNA ladder	Bioline
24- well plates	Molecular Dimensions®
2x YT broth	Thermo Fisher Scientific
96- well plates	Molecular Dimensions®
Acetic acid	Sigma- Aldrich
Acetonitrile	Sigma- Aldrich
Agarose	Sigma- Aldrich
All crystallography screens	Molecular Dimensions®
All mutagenic primers	Thermo Fisher Scientific
All peptides	Genscript
Ammonium molybdate	Sigma- Aldrich
Ampicillin	Sigma- Aldrich
Anti-mouse IgG alkaline phosphatase conjugate	Novagen®
APS	Sigma- Aldrich

Material	Supplier
Ascorbic acid	Sigma- Aldrich
BCA assay kit	Thermo Fisher Scientific
BCIP/NBT tablet	Sigma- Aldrich
BL21(DE3) cells	Agilent
BME	Sigma- Aldrich
Brilliant blue G-250	Thermo Fisher Scientific
Bromophenol blue dye	Sigma- Aldrich
$CaCl_2$	Sigma- Aldrich
Centrifugal concentrators	Sartorius
Chelating sepharose	Sigma- Aldrich
DH5α cells	Agilent
DNA loading dye	Thermo Fisher Scientific
DPNI	New England Biolabs
EDTA	Sigma- Aldrich
Ethanol	Sigma- Aldrich
Gel extraction kit	Qiagen
Gel filtration calibration kit	Sigma- Aldrich
Glutaraldehyde	Sigma- Aldrich
Glycerol	Sigma- Aldrich
Glycine	Sigma- Aldrich
HEPES	Sigma- Aldrich
IPTG	Generon
LB broth	Lennox
Lysozyme	Sigma- Aldrich
$MgCl_2$	Sigma- Aldrich
Miniprep kit	Qiagen
mouse anti-his antibody	Novagen®
NaCl	Sigma- Aldrich
$NiCl_2$	Sigma- Aldrich

Material	Supplier
Nitrocellulose membrane	Thermo Fisher Scientific
N- terminal signalling peptides (20 residue)	Generon®
PBS tablets	Sigma Aldrich
PD- 10 column	GE Healthcare
PEG 400	Molecular Dimensions®
PEG 600	Molecular Dimensions®
Pre- cast PAGE gels	Thermo Fisher Scientific
Protease inhibitor	Thermo Fisher Scientific
Q5 High- fidelity mastermix	New England Biolabs
SDS	Sigma Aldrich
Sinapinic acid	Sigma- Aldrich
Sulphuric acid	Sigma- Aldrich
SYBR® safe	Invitrogen
SYPRO™ orange	Thermo Fisher Scientific
TEMED	Sigma- Aldrich
TFA	Sigma Aldrich
Thrombin	Generon
Tris	Sigma- Aldrich
Tween- 20	Sigma- Aldrich
Uranyl acetate	Agar Scientific

2.2 Microbial and molecular biology techniques.

2.2.1 Site directed mutagenesis.

For site directed mutagenesis, a primer was designed to include a mismatched base pair(s) corresponding to the desired mutation using the Agilent Quick change primer design App. A summary of primers used can be found in table 2.2.

A solution containing: NEB Q5 High- fidelity 2X master mix (containing Q5 DNA polymerase, 2 mM Mg^{2+} and 200 μM of each deoxynucleotide), 1 μM forward primer, 1 μM reverse primer and 1 μg template DNA was made up to a final volume of 50 μl, using nuclease free water. 50 μl sample was split into five 10 μl aliquots, each in PCR tubes. The PCR proceeded as in table: 2.3. The reaction was performed across a temperature gradient, allowing for optimisation of the annealing temperature. Only the aliquot with the highest amplification (as assessed by agarose gel electrophoresis) was selected for transformation..

Following PCR, the restriction enzyme DPNI was added in order to digest any remaining methylated template DNA. The reaction was left to proceed overnight at room temperature before amplification was assessed using Agarose gel electrophoresis (section: 2.2.2).

TABLE 2.2: Summary of mutagenic primers used and the desired mutations at the protein level.

Protein	Desired mutation	Codon change	Primer
PduA	N67C	AAT >TGT	5′-cacttctcccacacaacgtgctgcggcggc-3′
			5′-gccgccgcagcacgttgtgtgggagaagtg-3
PduA	A63R	GCT >CGT	5′-cattacgtgctgcgcgggcacctgcatctg-3′
			5′-cagatgcaggtgcccgcgcagcacgtaatg-3′
PduK	A100STOP	GCC >TAG	5′-acgaccggtgtcggctagggcgctggctcag-3′
			5′-ctgagccagcgccctagccgacaccggtcgt-3′
PduL	D15C	GAC >TGC	5′-gctcacgcatttcacatcgaactttggtgaccgttgtctcca-3′
			5′-tggagacaacggtcaccaaagttcgatgtgaaatgcgtgagc-3′

TABLE 2.3: PCR protocol for site directed mutagenesis. Following Initial denaturation; denaturation, annealing and elongation were performed iteratively, 30 times, following one another, before the final elongation step.

Number of cycles	Step	Time (S)	Temp (°C)
1	Initial denaturation	30	98
30	Denaturation	10	98
30	Annealing	30	50- 72
30	Elongation	100	72
1	Final extension	120	2

2.2.2 Agarose gel electrophoresis.

A 1 % agarose gel was prepared by adding 1 g of agarose to 100 ml of TAE buffer (40 mM Tris pH 8.3, 20 mM Acetic acid, 1 mM EDTA). The mixture was heated, in 30 second pulses until all agarose had visibly dissolved. Once cool enough to comfortably hold $10\mu l$ of SYBR^{TM} safe was added, before the solution was poured into a gel tray and left to set.

The set gel was placed in a BioRad ReadySub GT cell and TAE buffer added until wells were submerged. 2 μl of 6 x DNA loading dye (thermofisher) was added to 10 μl of DNA sample in order to facilitate visualisation. This was pipetted into the wells, alongside a single well containing 5 μl Bioline 1kb Hyperladder in order to determine the size of amplified DNA.

2.2.3 Gel extraction.

Successful amplification of mutant DNA was verified by the presence of a band at the expected size, when examined under UV light. The band was excised using a scalpel and purified using the Qiagen QIAQuick gel extraction kit, as per the manufacturers instructions. The eluted DNA was then used to transform *DH5α* competent cells (section: 2.2.5) for plasmid amplification.

2.2.4 Preparation of chemically competent cells.

Either $BL21(DE3)$ or $DH5\alpha$ cells (NEB) were streaked onto LB- Agar plates under sterile conditions and grown overnight at 37 °C. A single colony was selected, then added to 200 ml sterile LB and grown to an OD of 0.2 at 37 °C. The media was transferred into chilled, sterile centrifuge tubes and the cells were harvested by centrifugation at 4000 xg (4 °C) for 10 minutes, before being placed on ice. Cells were then resuspended in 50 ml ice cold 100 mM $MgCl_2$ and left on ice for five minutes. The cells were separated through centrifugation at 4000 xg (4 °C) for 10 minutes and the supernatant discarded. The pellet was resuspended in 25 ml 100 mM ice cold $CaCl_2$, and left on ice for 20 minutes. The pellet was harvested by centrifugation at 4000 xg (4 °C) for 10 minutes. Finally, the pellet was resuspended in 2 ml 100 mM $CaCl_2$, 15 % glycerol and separated into 50μl aliquots, which were then flash frozen in liquid nitrogen and stored at -80 °C for future use.

2.2.5 Transformation.

1 μl of plasmid DNA was added to 50 μl of competent cells (section: 2.2.4) and left on ice for 40 minutes. Subsequently, cell were heat shocked at 42 °C for exactly one minute and placed back on ice for a further 5 minutes. 200 μl 2x YT was added and cells were incubated for one hour at 37 °C. The cell mixture was streaked out onto LB- Agar plates, containing 1mg/l ampicillin and grown overnight at 37 °C. If the transformed cells were $BL21(DE3)$ protein purification proceeded (section: 2.3.1), if they were $DH5\alpha$ for plasmid amplification DNA extraction followed(section: 2.2.6).

2.2.6 DNA extraction.

A singe colony of $DH5\alpha$ transformed cells were used to inocculate 10 ml LB- Agar, containing 1 μg/ml ampicillin. This culture was then grown at 37 °C overnight, before the pellet was isolated by centrifugation at 4000 xg for 10 minutes. Plasmid DNA was extracted from the pellet using the Qiagen spin miniprep kit, following the manufacturers instructions. Where required, sequencing was performed by eurofins overnight sequencing service and analysis undertaken using snap gene.

2.3 Protein purification and preparation.

2.3.1 Protein production.

All proteins were produced in BL21(DE3) cells. A single colony was selected and grown overnight at 37 °C, in a 10 ml 2x YT starter culture, containing 1 μg/ml ampicillin. 1 L of 2x YT containing 1 μg/ml ampicillin was inoculated with the starter culture and grown at 37 °C, until an OD of 1.0 was attained. Expression was induced by the addition of 1mM IPTG and cells were grown overnight at 18 °C, shaking at 200 RPM.

The 1L cultures were decanted into 500 ml centrifuge tubes and cells were harvested via centrifugation at 6000 xg for 20 minutes, at 4 °C. The resultant cell pellet was then resuspended in 50 ml lysis buffer (see table: 2.4). Lysis of the resuspended pellet was achieved using sonnication for 10 minutes at 20 W in 30 second on/off pulsing cycles. The cell lysate was clarified by centrifugation at 20,000x g for 30 minutes (4 °C).

2.3.2 Nickel affinity chromatography.

A nickel IMAC column was prepared by adding approximately 1 ml of chelating sepharose to an empty 10 ml column. The resin was charged by washing with 10 ml $NiSO_4^{2-}$ (50 mM) and then washed with several column volumes of water. The column was equilibrated with 15 ml of lysis buffer (table :2.4) before the sample was loaded. Samples were subsequently washed with increasing concentrations of imidazole, step- wise, with 10 ml of wash buffers (see table: 2.5). Before finally eluting with 10 ml elution buffer (see table: 2.5).

TABLE 2.4: Table of the constituents of buffers used in the lysis of cell pellets. All lysis buffers contain 25 mM imidazole and 500mM NaCl. PduA (K26D) and $PduA_{GST}$ also refers to the concatenated variants of these proteins.

Protein(s)	Buffer (pH)	Glycerol(%)	BME (mM)	protease inhibitor	Lysozyme
PduA*	TRIS (8.5)	0	0	yes	10mg
$PduA_{GST}$	TRIS (8.5)	0	0	yes	10mg
PduA (K26D)	TRIS (8.5)	0	0	yes	10mg
PduA (A63R)	TRIS (9.5)	5	0	yes	10mg
PduA (N67C)	TRIS (8.5)	0	1	yes	10mg
PduB	TRIS (8.5)	0	0	yes	10mg
PduK	HEPES (7.5)	10	1	yes	10mg
PduL	HEPES (8.0)	10	0	yes	10mg
PduL (D15C)	HEPES (8.0)	10	1	yes	10mg
PduP	HEPES (8.0)	10	0	yes	10mg

TABLE 2.5: Table showing the imidazole concentration in buffers used for the purification of proteins. All other buffer constituents are identical to those in table: 2.4. PduA (K26D) and $PduA_{GST}$ also refers to the concatenated variants of these proteins.

	Imidazole concentration (mM)				
Protein	Equilibration	Wash 1	Wash 2	Wash 3	Elution
PduA*	25	25	50	100	500
$PduA_{GST}$	25	25	50	100	500
PduA (K26D)	25	25	50	100	500
PduA (A63R)	25	25	50	100	500
PduA (N67C)	25	25	50	100	500
PduB	25	25	50	100	500
PduK	25	25	50	100	500
PduL	25	25	50	80	150
PduL (D15C)	25	25	50	80	150
PduP	25	25	40	60	150

2.3.3 Buffer exchange of target proteins.

The fraction eluted following nickel was desalted using a PD- 10 column (GE Healthcare). The column was pre- equilibrated with 20 ml storage buffer (which was identical to the elution buffer with the absence of imidazole). 2.5 ml of the protein sample was loaded onto the column, discarding any flow through. 3.5ml of storage buffer was added and the desalted protein elutant was collected.

2.3.4 Cleavage of hexahistidine tag.

For enzymes with N- terminal signalling sequences it was necessary to cleave the his- tag so as to not interfere with potential interactions. The pET14b vector features a thrombin cleavage sequence (Leu-Val-Pro-Arg-Gly-Ser) upstream of the protein and downstream of the his- tag, allowing for removal of the latter (figure: 2.1). Proteins were transferred to PBS buffer, using a PD- 10 column (section:2.3.3 and thrombin was added to a final concentration of 50 units/ml. The reaction was incubated at 4, 18 and 30 °C for 1, 6 or 12 hours. Cleavage efficacy was determined using SDS- PAGE. Where optimal cleavage had occurred, the sample was loaded onto a nickel IMAC column and washed with PBS. Any protein where the his- tag has been successfully cleaved will not bind to the charged resin and this flow through was collected.

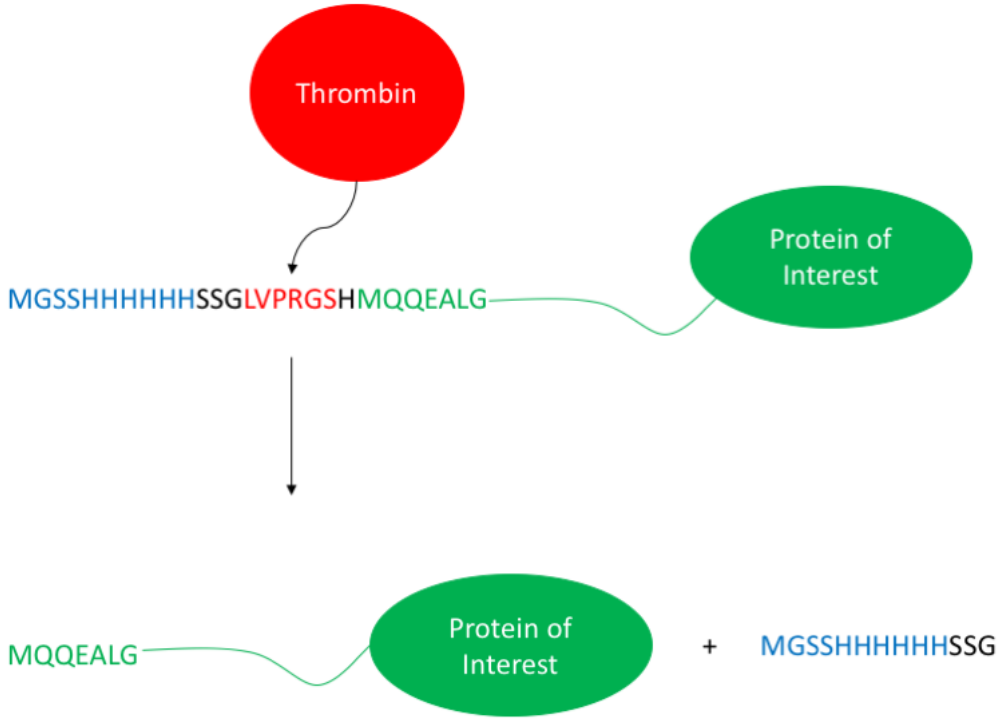

FIGURE 2.1: Diagramatic representation of thrombin cleavage. Showing the his- tag (blue), thrombin consensus sequence (red) and protein of interest (green). Thrombin acts to cut at the consensus sequence, which is upstream from the protein of interest and downstream of the his- tag, resulting in the removal of the his- tag from the protein of interest.

2.3.5 Size exclusion chromatography.

Protein was concentrated to 1ml and loaded onto a Superdex 200 Increase 10/300 GL column, pre- equilibrated with storage buffer. Eluted fractions containing protein were identified according to a UV chromatogram and further analysed using SDS-PAGE. Fractions containing pure protein were pooled and used either immediately for experiments or flash frozen in liquid nitrogen and stored at -80 ° C for later use.

2.4 Verification of protein identity and concentration.

2.4.1 SDS- PAGE.

SDS- PAGE was performed using a 4- 15 % precast gradient gel (Thermofisher) unless otherwise stated. Where fixed percentage gels were used, the separating gel was prepared by adding acrylamide, ddH_2O, Tris- HCl, TEMED, APS and SDS in the proportions shown in table 2.6 to a volume of 10 ml. This mixture was immediately transferred into a 1 mm thick gel mould and 200μl of isopropanol was added on top, to keep the surface of the gel even and evaporate whilst the gel set. A stacking gel was prepared as shown in table 2.6 and added to the top of the separating gel. A 1 mm 12- well comb was quickly inserted and the gel allowed to set.

Protein samples were boiled in loading buffer (10 % SDS, 500 Mm BME, 50 % Glycerol, 500 mM Tris- HCl and 0.05 % bromophenol blue dye) for 5 minutes before loading into the wells. Electrophoresis was performed at 200V in running buffer (192 mM glycine, 25 mM Tris pH 8.5, 0.1 % SDS), until the dye front had reached the end of the gel.

The gel was soaked in coomassie stain (0.1 % coomassie brilliant blue G- 250, 30 % ethanol, 10 % Acetic acid, 60 % H_2O) for 20 minutes.The gel was then washed in distilled water and de- stained in de- staining solution (30 % ethanol, 10 % Acetic acid, 60 % H2O) until the protein bands were easily identifiable by eye. Images of the gel were then taken using the Bio -Rad Gel Doc imaging system.

TABLE 2.6: Table showing the constituents of acrylamide gels, where pre- cast gels were not used. Where gels were used for SDS- PAGE 100 μl 10 % SDS was added.

Percentage	Acrylamide (30 %)(ml)	Tris pH 8.5 (1.5 M)(ml)	Tris pH 6.5 (0.5 M)(ml)	ddH2O (ml)	TEMED (μl)	APS (μl)
4 (stacking)	0	0.85	1.5	2.5	15	50
8	2.75	3	0	4	15	100
10	3.3	3	0	3.3	15	100
12	4	3	0	2.7	15	100

2.4.2 Concentration calculation using A_{280}.

The Beer- Lambert law (see equation: 2.1) was used as the primary method for determining protein concentrations. A is the absorbance at 280 nm of the protein samples, ε is the extinction coefficient of the protein, C is the concentration of the protein sample and l is the path length of the cuvette. The extinction coefficient was estimated using the ExPasy ProtParam software [62].

$$A = \varepsilon c l \tag{2.1}$$

2.4.3 Concentration calculation using BCA Assay.

Proteins which do not contain Trp or Tyr residues do not allow for reliable estimations of protein concentration. Therefore, BCA assays were performed where proteins lacked these residues.

Standards of BSA were prepared in concentrations of 0.03, 0.06, 0.125, 0.25, 0.5, 1.0 and 2.0 mg/ml. 20 ml of the Pierce BCA assay reagent A was added to $400\mu l$ of reagent B and mixed thoroughly. 10 μl of each of the BSA standards was added to 200 μl of this mixture and incubated together at 37 °C for 30 minutes. The A_{562} was measured and the data plotted on a standard curve (see figure: 2.2). For samples of unknown concentration, 10 μl of sample was incubated with 200 μl of premixed reagents. The standard curve was then used to derive protein concentration.

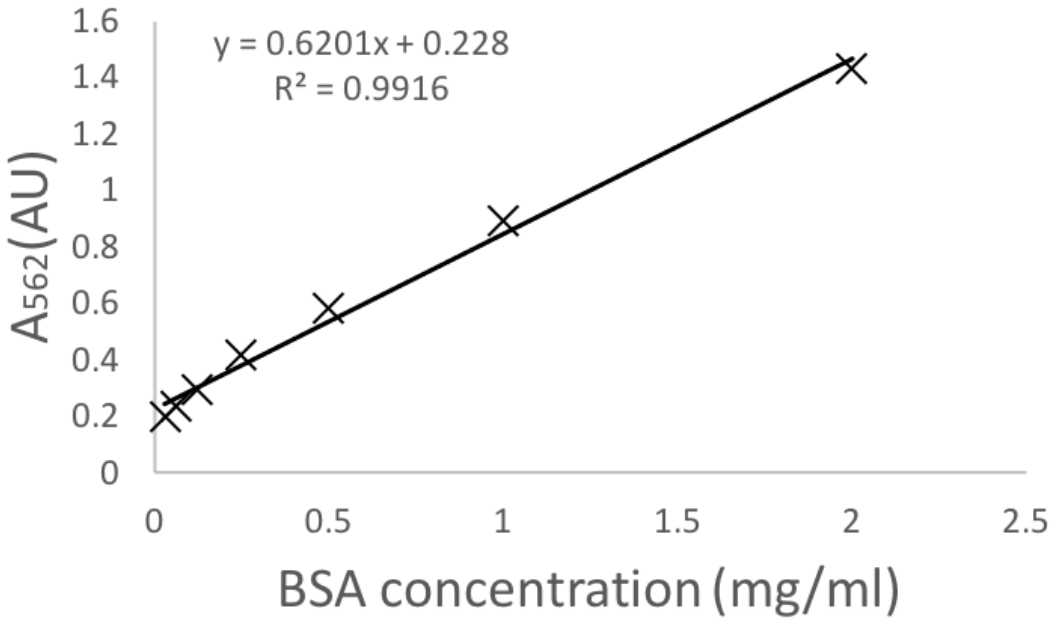

FIGURE 2.2: BCA Assay standard curve, showing known concentrations of BSA plotted against their corresponding A_{562} values.

2.4.4 Calibration of columns.

Superdex 200 Increase 10/300 GL and Superose 6 10/300 GL ™ columns were equilibrated in order to approximate the size of purified proteins. The Sigma Aldrich gel filtration calibration kit was used, containing five different proteins of varying molecular weights. These proteins were dissolved together in 1 ml of water, each to a final concentration of 1 mg/ml. The protein mixture was loaded onto the column and run for a full column volume. The $K_a v$ was calculated from the elution profile using equation 2.2. Where V_e is the volume at which the protein elutes, V_0 is the void volume of the column and V_t is the total column volume.

$$K_a v = \frac{V_e - V_0}{V_t - V_0} \tag{2.2}$$

The $K_a v$ was then plotted against logMW and this standard curve could be used to estimate the molecular weight of other proteins from the elution volume.

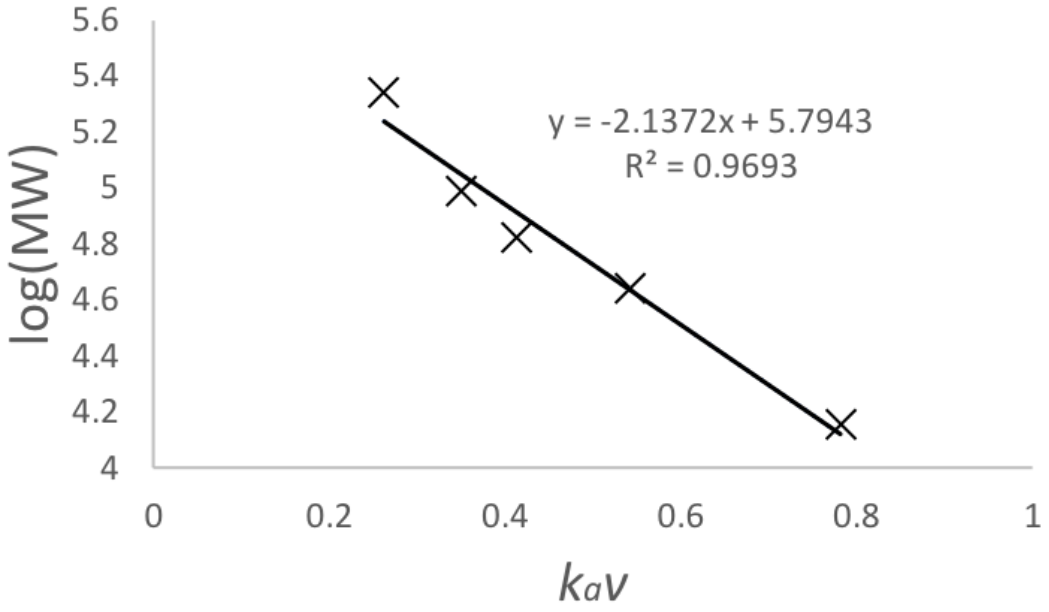

FIGURE 2.3: Standard curve for an Superdex S200 size exclusion column, the $K_a v$ is calculated from the elution volume of supplied standard proteins and plotted against the log of the known molecular weight

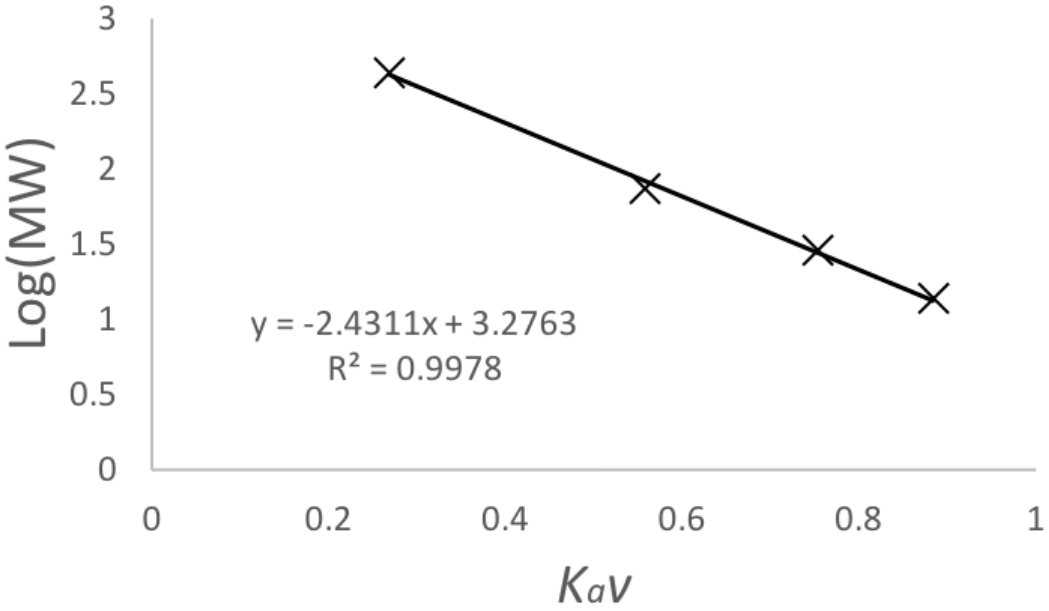

FIGURE 2.4: Standard curve for an Superose 6 size exclusion column, the $K_a v$ is calculated from the elution volume of supplied standard proteins and plotted against the log of the known molecular weight

2.5 Assessing the oligomeric state of PduA.

2.5.1 Glutaraldehyde cross- linking.

A 1 ml volume of PduA, at a concentration of 1 mg/ml was sonicated briefly for 30 seconds in an ice bath, to disrupt any larger structures that had formed. Glutaraldehyde was added to a final concentration of 0.025 % and the reaction was left to proceed. Every 30 seconds, 100 μl was removed and Tris pH 8.0 was added to a final concentration of 100 mM in order to terminate the reaction. The samples were then assessed using SDS- PAGE to determine the optimal timescale for glutaraldehyde treatment. A time range was identified that allowed for detection of larger PduA oligomers, whilst ensuring the majority was not present as large aggregates which could not diffuse through the acrylamide matrix.

The experiment was repeated for a suitable timescale and the sizes of species produced were assessed using a Superdex S200 column.

2.5.2 Native PAGE.

Native PAGE was used to assess both the oligomeric state of PduA and related mutants and also to investigate the interaction between PduA and PduL.

Native PAGE gels used were either 4- 15 % acrylamide precast gradient gels, or where stated fixed percentage acrylamide gels. The constituents of these gels can be found in table:2.6. Native PAGE was performed using a buffer containing 192 mM glycine and 25 mM Tris pH 8.5, the samples were run for 2 hours at 100V in a cold room (4 °C). Where proteins analysed had a pI greater than the pH of the running buffer (8.5) an identical running buffer with a pH of 9.5 was used. This was necessary as there is no SDS used in native PAGE, which functions to impart a negative charge on unfolded proteins. Therefore, proteins migrate entirely based upon their intrinsic negative charge in native PAGE and so the pH of the buffer must be greater than the pI of the protein in order to achieve a net negative charge.

Completed Native PAGE gels were stained and imaged in a manner identical to SDS-PAGE (see section: 2.4.1).

2.5.3 Western Blot.

In certain Native PAGE experiments it was necessary to demonstrate it was PduA migrating further and not just PduL giving that appearance. The his- tag of PduL had been cleaved (section: 2.3.4) to facilitate binding to PduA, enabling the exclusive detection of PduA with an anti- his antibody.

Two Native PAGE gels were run in parallel for 2 hours. One was stained in the manner described in section 2.4.1, and the other was transferred onto a nitrocellulose membrane for western blotting.

Transfer to the nitrocellulose membrane was achieved using the Bio- Rad Trans-Blot SD Semi-Dry Transfer Cell. Filter paper, nitrocellulose membrane and gel were soaked in cold transfer buffer (table: 2.7) for one hour, before components were arranged as in figure: 2.5 and transfer proceeded for one hour at one amp. It is imperative that SDS is used in the transfer buffer where native gels are blotted as the SDS is required at this point to impart the charge required for successful transfer. Non- specific binding was prevented by soaking in blocking buffer (table: 2.7), overnight at room temperature. The membrane was washed three times in TBST (table: 2.7) and the primary antibody (mouse anti-his) (Novagen®) and membrane were incubated together overnight at 4 °C. The membrane was washed three times in TBST and incubated with the secondary antibody (anti-mouse IgG alkaline phosphatase conjugate) (Novagen®) for a further hour at room temperature. The membrane was washed three times in TBST and colour was developed using SIGMAFAST BCIP/NBT tablet dissolved in water. Here, the alkaline phosphorylase conjugated to the secondary antibody hydrolyses BCIP which is oxidised by NBT. This oxidised species has an intense purple colour, enabling visualisation.

TABLE 2.7: A table showing the composition of buffers used for western blotting

Buffer	Composition
Transfer	25 mM Tris, 192 mM glycine, pH 8.5, 20 % methanol
Blocking	1X TBST, 5 % milk powder
TBST	20 mM Tris, 150 mM NaCl, 0.1% Tween 20, pH 7.4

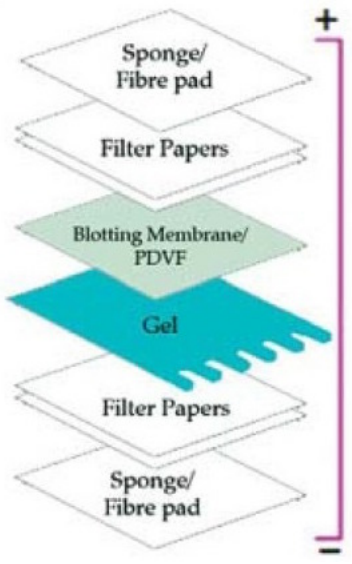

FIGURE 2.5: Diagramatic representation of the filter paper, nitrocellulose membrane and gel and how they are arranged in the transfer cell in respect to the cathode and anode. Adapted from Mahmood (2012) [63]

2.5.4 MALDI- TOF.

Matrix assisted laser desorption/ ionisation- Time of flight (MALDI- TOF) was performed using a bruker rapifleX MALDI PharmaPulse spectrometer. MALDI- TOF works on the principles that a charge is imparted on the molecule of interest from a laser, via a matrix which is able to absorb laser radiation and donate protons to the analyte. The ionised particle travels through a known electric field (V) where a detector registers the impact and provides a value for time of flight (t). As the length distance between the matrix and detector is also known (d), the mass to charge ratio (m/Q) can be derived from equations 2.3 Typically, the charge on a species generated from MALDI is +1 and so the mass is equal to the mass: charge ratio.

$$E = QV$$

$$E = \frac{(mv^2)}{2}$$

Therefore:

$$\frac{(mv^2)}{2} = QV$$

(2.3)

Sinapinic acid was used as the matrix for MALDI experiments described here, and is a regarded as a one of the more appropriate matrices for use with larger (<10kDa) molecules [64]. Sinapinic acid was added in excess to 30 % H_2O, 70 %Acetonitrile, 0.1 % TFA. This mixture was then vortexed thoroughly and centrifuged for 1 minute at 20,000RPM. The supernatant (saturated with sinapinic acid) was removed and added to 1 mg/ml of desalted protein (section: 2.3.3) in a 1:1 ratio. 1 μl of this mixture was added to the sample plate and left for 30 mins or until thoroughly dried. Sample analysis then proceeded using positive detection mode, collecting 1000 images per sample.

2.6 Assessing the interaction between PduA and PduL.

2.6.1 Thermal melt experiments.

A qPCR machine and SYPRO® orange dye were used to generate thermal melt profiles of individual proteins and protein complexes. SYPRO® orange is a fluorescent dye which binds to hydrophobic regions of proteins [65]. These regions are typically buried in the proteins native state but as the protein unfolds, as a function of increasing temperature, the dye is able to bind to these regions and a fluorescent signal is recorded. For all thermal melt assays, protein was added to a final concentration of 1 mg/ml in a volume of 20 μl in a 96 well plate, to which 5 μl 5x SYPRO® orange was added. The fluorescence was then monitored over a temperature range of 20-95 °C, with temperature increasing by 1 °C every minute.

2.6.2 Crystallography.

Theory.

X- ray crystallography is a powerful technique for analysing the atomic structures of macromolecules and can give precise information about the atomic co- ordinates. From these co- ordinates, further inferences can be made about the overall structure and function of the macromolecule. Crystallography relies on the highly ordered, highly symmetrical arrangement of atoms in a crystal lattice. When a crystal is irradiated, assuming the radiation source has comparable wavelength to the distance between atoms in the crystal, a diffraction pattern will be produced. The distance (d), and therefore the atomic co- ordinates can then be calculated from the diffraction pattern using Braggs law: 2.4. Here, d is the distance between parallel planes in the crystal lattice, θ is the angle of incidence, λ is the wavelength of the incident radiation and n is the difference in path length of reflected waves (expressed in numbers of wavelength). Where n is a whole number, the reflected waves will interfere constructively, producing an increase in intensity and a well defined diffraction pattern. However, where n is a fraction the reflected waves will interfere destructively, leading to a decrease in intensity and a poorly defined diffraction pattern. As a

diffraction pattern will only be produced at specific values for θ where n is an integer, the unknown value d can be calculated. A simplified diagram, showing only two planes is shown in figure 2.6 [66].

$$2dsin\theta = n\lambda \tag{2.4}$$

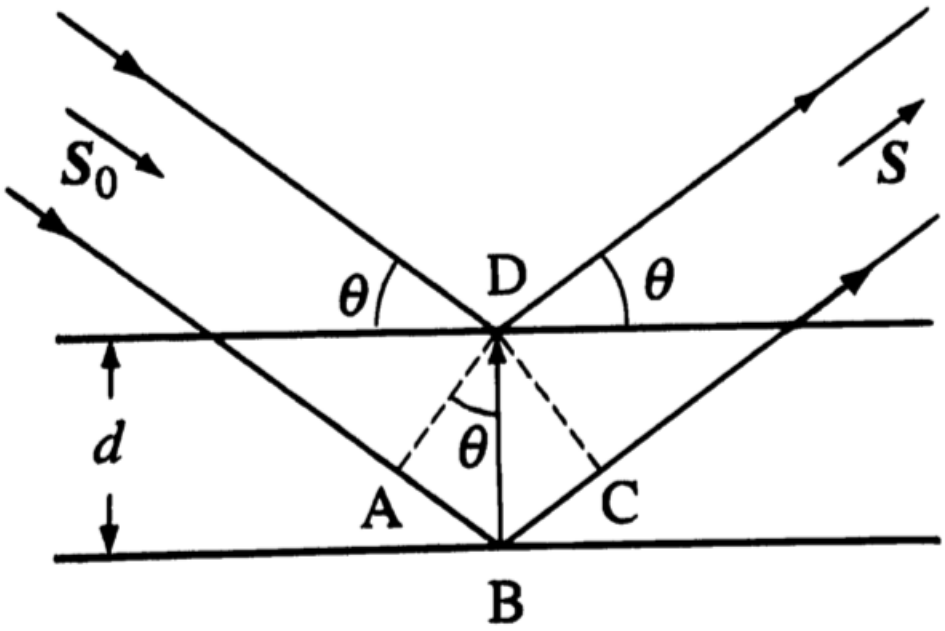

FIGURE 2.6: Diagramatic representation of Bragg diffraction. An incident wavelength (S_0) reflects from either point D and point B. The difference in path length of reflected waves (S) is equal to point A to B + B to C. Where this difference is equal to a whole number of wavelengths of S_0 the reflected waves (S) will interfere constructively and d can be derived. Figure adapted from Fundamentals of Crystallography (Giacovazzo, 2015) [66].

Crystallography screening and optimisation.

Purified protein (section: 2.3.5) was concentrated to 10 mg/ml using a centrifugal concentrator (sartorius) and verified using a BCA assay (section: 2.4.3). Initial

screening was performed in 96- well plates, using the hanging drop vapour diffusion method. Plates were prepared using 200 nl of the following commercially available screening sets: Structure Screen 1+2 HT-96 (MD1-30), JCSG-plus HT- 96 (MD1-40), MemGold 2 (MD1-63), PACT premier (HT-96) and Morpheus HT-96 (MD1-47). All screening kits were provided by Molecular Dimensions®. Initial screens were prepared using the TTPLabTech Mosquito, with 200 μl each of protein solution and screening condition added in a single drop. Trays were sealed and left at 19 °C. These trays were examined for crystal formation by light microscopy in the subsequent weeks. Where evidence of crystal formation was apparent, these conditions were optimized further to produce larger, more useful crystals. To achieve this, slight variations were made to the conditions yielding crystals and trialed in a 24- well plate with 1 ml of solution in the reservoir and 1 μl of each protein solution and well solution added to a cover slide in a single drop. This cover slide was sealed on top of the well using vacuum grease. The tray was stored at 19 °C and the formation of crystals was monitored.

Crystal picking, diffraction data collection and processing.

Suitable crystals, of an appropriate size were picked using Mounted Round LithoLoops (Molecular Dimensions®), with the assistance of a light microscope. Where the crystallisation solution did not contain a suitable cryo- protectant (glycerol, PEG400/600), the crystals were briefly soaked in a cryo- solution composed of the mother liquor with additional glycerol to a concentration of 20 %. Mounted crystals were transferred to a Uni-Puck, immersed in liquid nitrogen, and transferred to a pre- cooled dewar. The dewar was shipped to Diamond Light Source via the crystallography department at Birkbeck, University of London. Diffraction data was collected using GDA (Generic Data Acquisition) software remotely through the use of the No-Machine. Three images of each crystal were taken at 45 ° oscillations in order to generate a refinement strategy using Xia2/3dii [67].This strategy was used to collect complete dataset composed of many diffraction images, 1800 in the case of PduL. Diffraction data were reduced using Mosflm [68] and phases were calculated using the molecular replacement software, Phaser in the Phenix suite [69]. hlThe search

models used for molecular replacement were PDB: 5CUO and PDB: 3N79 for PduL and PduT respectively. Refinement of the model was achieved through iterative rebuilding in Coot [70] and use of the refinement tool in Phenix to generate values regarding the quality of the model. Initial refinement following molecular replacement was performed using rigid body refinement for 3 macrocycles. All subsequent refinement cycles used restrained ADP refinement and occupancy refinement for three macrocycles. Analysis of the validity of oligomeric states within the crystal lattice was undertaken using the PISA webserver [71] and final validation performed using MolProbity [72]. All subsequent images generated of structural data were created using PyMOL [73]

2.6.3 TEM.

Transmission Electron Microscopy (TEM) works on the principle that high energy electrons will scatter as they pass through a thin (<100nm) sample [74]. The amount of scattering is dependant on the density and depth of the sample, and areas of very high density are impermeable to electrons. The electrons which pass through the sample are focused onto a phosphor screen and magnified, allowing for imaging of the sample. This technique is particularly useful for analysing large, macro- structures formed by proteins. Especially where these structures are too heterogenous to allow for crystallography, but too small to be imaged by conventional light microscopy.

For preparation of TEM samples, a solution of 1 % (w/v) uranyl acetate was prepared, in distilled water. Square mesh copper TEM grids (Labtech) were treated in a glow discharger. This process ionises gas in the local environment, depositing carbon on the grid in the process and making it more hydrophilic. The grid was held in place using negative pressure tweezers, to avoid damage. 5 μl of protein sample was applied and allowed to soak in for 90 seconds. Excess sample was removed by blotting with filter paper and the grid washed with 5 μl ddH_2O, with excess water removed using filter paper. Uranyl acetate was applied in a 5 μl quantity and left to stain the sample for 70 seconds. The 1 % uranyl acetate solution was always filtered using a 45 micron filter immediately before use, in order to remove any large

molecules of precipitate. The grid was washed with 5 μl ddH_2O and dried with filter paper. Grids were loaded onto a Jeol JEM-1230 TEM and analysed before final images were taken using an Olympus Morada 2K camera.

2.6.4 CD.

Circular dichroism (CD) is a useful tool for examining the secondary structure of proteins. Whilst no detailed structural information is discernible, changes to secondary structure following binding events can be detected and used to infer interactions between proteins. CD uses left and right circularly polarised light, which is absorbed in a differential manner by peptide bonds depending on the phi and psi angles. These bond angles are heavily influenced by the secondary structure of the protein, and so characteristic spectra are produced depending on the dominant secondary structure of the protein analysed [75]. Where proteins are predominantly alpha helical, peak absorbance at 208 and 222nm is observed. Where a structure is predominantly $_{b}eta$- sheets or disordered regions these peaks are at 216 and 195nm respectively.

All CD samples were prepared in a buffer of 25 mM PO_4^{2-}. The concentrations of PduA and signalling peptides used were 0.5 and 4 mg/ml respectively. For assays investigating the binding of PduL to PduA, the PduL signalling peptide (L20) was added to PduA* to give a 6:1 (peptide: hexamer) molar ratio. Circular dichroism was measured using a Chirascan^{TM} spectrometer, in a 0.05mm path length cuvette for all samples. Molar ellipticity was calculated using equation: 2.5 and CD spectra were analysed using the K2D3 software [76] to estimate percentage helicity.

$$MolarEllipticity = \frac{mdeg}{33,000 * ProteinConcentration(M) * PathLength(cm)} \tag{2.5}$$

2.6.5 Fluorescence quenching assays.

Fluorescence quenching techniques rely on calculating protein binding parameters by quenching the intrinsic fluorescence of tryptophan. Tryptophan is excited at 295 nm and emits a fluorescence spectra with a peak at 355 nm. Where a ligand reduces the accessibility to tryptophan, be this through a direct interaction or merely as a result of restricting access, the fluorescence will decrease. By titrating a ligand into the protein sample, until a plateau is achieved (this occurs where the maxima number of ligands have bound) the association constant of the ligand can be calculated using equation 2.6. Given that the concentration of protein ([P]) and ligand ([L]) are known, the concentration of the complex ([PL]) can be calculated from the saturation curve. Allowing for the calculation of the dissociation constant (K_d).

All fluorescence quenching experiments were completed using a Hitachi High- tech F2700 fluorescence spectrophotometer and quartz cuvette with 1 cm path length. Samples were pre- prepared with 1 uM PduK and PduL in ratios ranging from 1:0-1:10 and spectra were recorded with an excitation wavelength of 295 nm and emission maxima at 355 nm.

$$K_a = \frac{[PL]}{[P][L]}$$

$$K_d = \frac{1}{K_a}$$

(2.6)

2.7 Kinetics experiments.

2.7.1 Phosphate assay.

For experiments pertaining to the activity of PduL, the loss of phosphate was measured using a molybdenum blue assay [77]. This technique relies on the strongly blue coloured complex, ammonium phosphomolybdate formed by phosphate and

ammonium molybdate in the presence of a reducing agent. This brightly coloured species absorbs strongly in the 820 nm region. The assay reagent (reagent C) was prepared using 1.2 M Sulphric acid, 0.5 % ammonium molybdate and 2 % ascorbic acid, in deionised water. A standard curve was generated using 0.1, 1, 10, 50 and 100 mM sodium phosphate. Here, 1ml of each of these known samples was added to 1 ml of reagent C and incubated at 37 °C for 1.5 hours. Absorbance at 820 nm was then recorded and a standard curve plotted using excel. For measurements of phosphate used by PduL, the protein was placed in a buffer containing 100 mM phosphate, 1 mM zinc acetate, 1 mM CoA and 100 mM HEPES (pH 7.5) and the reaction allowed to proceed for 10 minutes before terminating with the addition of 1 ml regent C. The mixture was incubated at 37 °C for 1.5 hours and the A_{820} was measured. The phosphate concentration present in the sample was then calculated using the standard curve.

2.7.2 CoA assay to determine substrate transfer and PduP activity.

PduP catalyses the conversion of propionaldehyde and CoA into propionyl- CoA, reducing NAD+ into NADH in the process. NADH absorbs strongly at 340 nm and was used to monitor the activity of PduP. Assays to determine the activity of PduP were constructed as has been previously described [78] using a reaction mixture containing 0.5 mM CoA, 5 mM NAD+, 100 mM propionaldehyde and 100 mM HEPES pH 8.0. Where PduL was also involved in the assay, this was supplemented with 100 mM sodium phosphate. The absorbance at 340 nm was recorded over a 1800 seconds and used to produced a graph of NADH concentration (calculated using the extinction coefficient of 6220 for NADH) against time. The gradient was taken from the initial portion of the graph and used to estimate initial rate. The specific activity was calculated by dividing the initial rate by the known protein concentration (in mg/ml).

2.8 Computational techniques.

2.8.1 Sequence alignments and analysis.

All sequence alignments were performed with the Clustal Omega multiple sequence alignments webserver [79] (https://www.ebi.ac.uk/Tools/msa/clustalo/).

2.8.2 Peptide modelling.

Peptide structures used in interaction modelling were predicted using the Phyre2 protein fold recognition webserver [80] (sbg.bio.ic.ac.uk/ phyre2/).

2.8.3 Prediction of protein and peptide Complexes.

For modelling the interactions between PduA and the signalling peptides a variety of docking softwares were trialled, shown in table: 2.8. The PduA model used in all docking simulations was PduA (PDB: 3NGK). Trimers of hexamers, generated from the symmetry mates of the PDB file were used in most cases. In the case of CABS- DOCK and HPEPDOCK there is a receptor residue limit of 600 residues and so only the six residues forming the interface at the centre of the trimer were used (figure: 2.7).

TABLE 2.8: Docking software used for insilico interaction studies. The name and address are shown as well as the basis for simulation.

Docking software	Basis for simulation	Address
CABS- DOCK	flexible protein: peptide docking	biocomp.chem.uw.edu.pl/CABSdock [81]
HPEPDOCK	blind peptide–protein docking based on a hierarchical algorithm	huanglab.phys.hust.edu.cn/hpepdock [82]
GalaxyPepDock	Protein-peptide docking based on interaction similarity	galaxy.seoklab.org [83]
flexpepdock	high-resolution peptide-protein docking refinement tool	flexpepdock.furmanlab.cs.huji.ac.il [84]

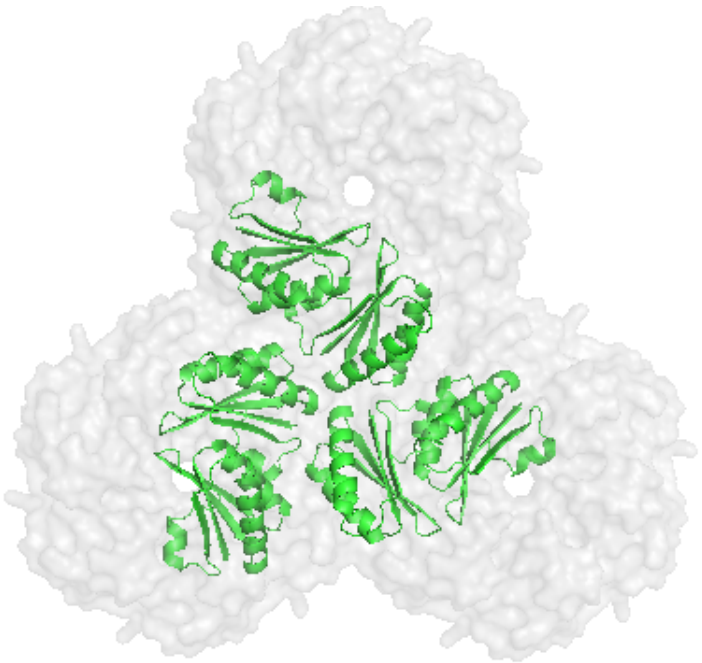

FIGURE 2.7: The PduA model used for docking simulations as a whole shown in transparent grey, with the smaller variant used for both CABS- DOCK and HPEPDOCK simulations shown in green.

Chapter 3

Tessellation of Pdu shell proteins.

3.1 Introduction:

Knowledge of BMC shell protein tessellation is fundamental in understanding the assembly of whole BMCs. Whether a shell protein preferentially interacts with itself or other shell proteins underpins the formation of the entire BMC, influencing the size of facets and the angles between them. The factors which contribute to tessellation have been investigated in recent years. Several mutagenic studies have demonstrated the importance of residues lysine 26, valine 51 and arginine 79 in the formation of PduA sheets [25]. These residues are conserved throughout the majority of shell proteins and are assumed to be essential in tessellation.

Recently the structure of an entire BMC from *H. ochraceum* [85] provided an insight into the interactions between different shell proteins. This demonstrated a preferential interaction between trimeric PduT- like proteins and hexameric PduA- like proteins, which are also capable of interacting with themselves. The hexameric proteins are oriented 180º relative to one another, forming the facets of the BMC. In contrast, the trimeric proteins interact exclusively with hexameric shell proteins forming a 150º angle at the interface, thereby facilitating the curvature of the compartment.

Whilst entire microcompartments and single protein nanostructures are detectable with TEM and crystallography, there is limited information available on how these structures assemble. It is unclear whether shell proteins exist as single hexamers in solution before the formation of higher order structures, or as some larger oligomers

which then themselves multimerise. In this chapter, the tessellation of the hexameric shell proteins, PduA and PduK is investigated using a variety of techniques (including chemical ligation and mutagenesis) in order to trap and characterise early assembly intermediates.

A variant of PduA featuring a 23- residue C- terminal extension, which enhances solubility has previously been described [25]. This variant, dubbed PduA* tessellates into tubes and sheets in the same manner as PduA, and is used in place of PduA in the subsequent chapters. For consistency and ease of comparison, PduA* will be referred to as PduA. All PduA mutants used also feature this 23- residue extension, unless explicitly stated otherwise.

3.2 Purification of PduA

His- tagged PduA was purified by nickel IMAC (section: 2.3.5). PduA was washed with buffers containing imidazole up to 100 mM. This is necessary as PduA has a tendency to co- elute with other protein contaminants unless washed with sufficiently high concentrations of imidazole. Fortunately, the hexameric nature of PduA allows for these high imidazole washes as each PduA hexamer features six his- tags and so will not elute at imidazole concentrations < 300 mM. PduA was eluted by a buffer containing 500 mM imidazole and the eluted fractions were analysed by SDS- PAGE. Fractions containing pure PduA were pooled and purified further by SEC. PduA eluted at 14.1 ml, a size approximately consistent with a single hexamer when compared to the calibration (section: 2.4.4). The fractions around the peak (13, 14 and 15 ml) were pooled and assessed for purity using SDS- PAGE (figure: 3.1), showing the protein had been purified to near homogeneity. This protein was then used immediately, or flash frozen in liquid nitrogen and stored at -80° C until required.

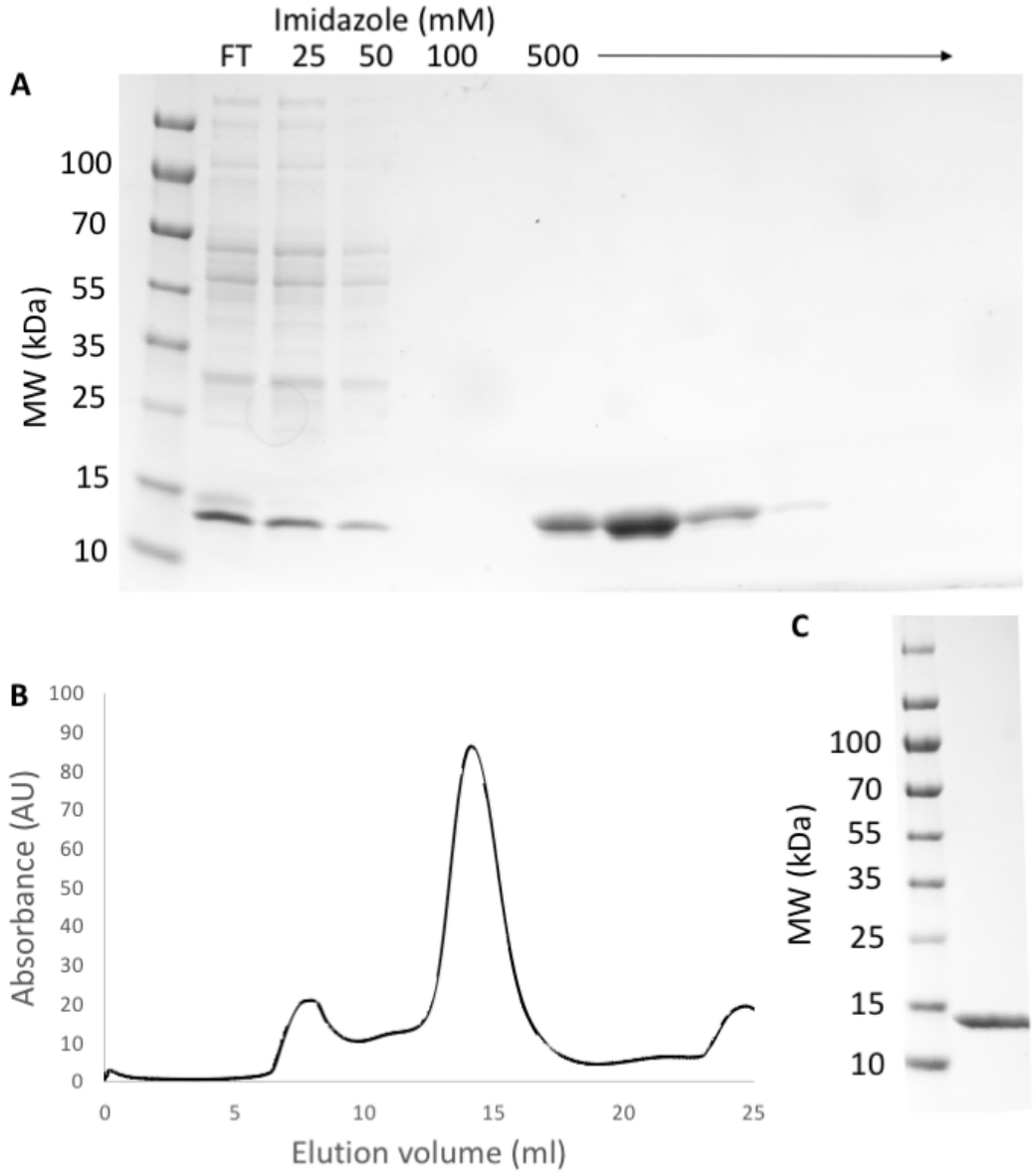

FIGURE 3.1: Typical purification of PduA. Showing the SDS- PAGE of PduA purified by IMAC (A). FT corresponds to the initial flow through following loading. The concentration of imidazole washes is shown above, before elution at 500 mM imidazole. The elution profile of PduA (B), following SEC using a Superdex S200 column shows PduA elutes at 14.1 ml, consistent with a single hexamer. SDS-PAGE of the final product following IMAC and SEC (C), shows PduA has been purified into a highly homogenous sample.

3.3 Investigating the tessellation of PduA.

Tessellation of PduA into tubes and sheets has been reported numerous times [18, 25, 30]. Typically, when PduA is concentrated in buffers containing low concentrations of NaCl (50 mM), tubes form and are visible when viewed using TEM [25]. The formation of PduA nanotubes can also be observed on a macroscopic level, where the appearance of a solution containing PduA changes from clear to a white, gel like precipitate (as the insoluble tubes form). Tubes with similar morphologies have also been reported in vivo, in cells over expressing PduA [18].

In the crystal structure of PduA, and of most shell proteins, the protein multimers assemble into flat sheets [24] (figure: 3.2). These sheets maintain the six- fold symmetry of the single hexamers by interacting with another hexamer in an identical manner at each of the six edges. The oligomeric state of PduA hexamers, before reaching a concentration threshold which drives the oligomerisation into tubes and sheets is poorly understood. Here, data are presented suggesting the existence of PduA as multimers of hexamers in solution, and at far lower concentrations than that required for the formation of tubes.

As previously mentioned lysine 26 is critical in the process of shell protein tessellation and is necessary to facilitate the packing of PduA in to flat sheets (figure: 3.2). A mutant of PduA, K26D, was shown to have a markedly reduced ability to tesselate, as assessed through TEM and crystallography (PDB: 4QIE) [25]. In addition to this mutant, a PduA variant with the extension (gly, gly, ser, gly, thr) *5 (instead of the 23 residue A* variant) was produced by Dr. Ruth Rose. This mutant is referred to as PduA_{GST} hereafter.

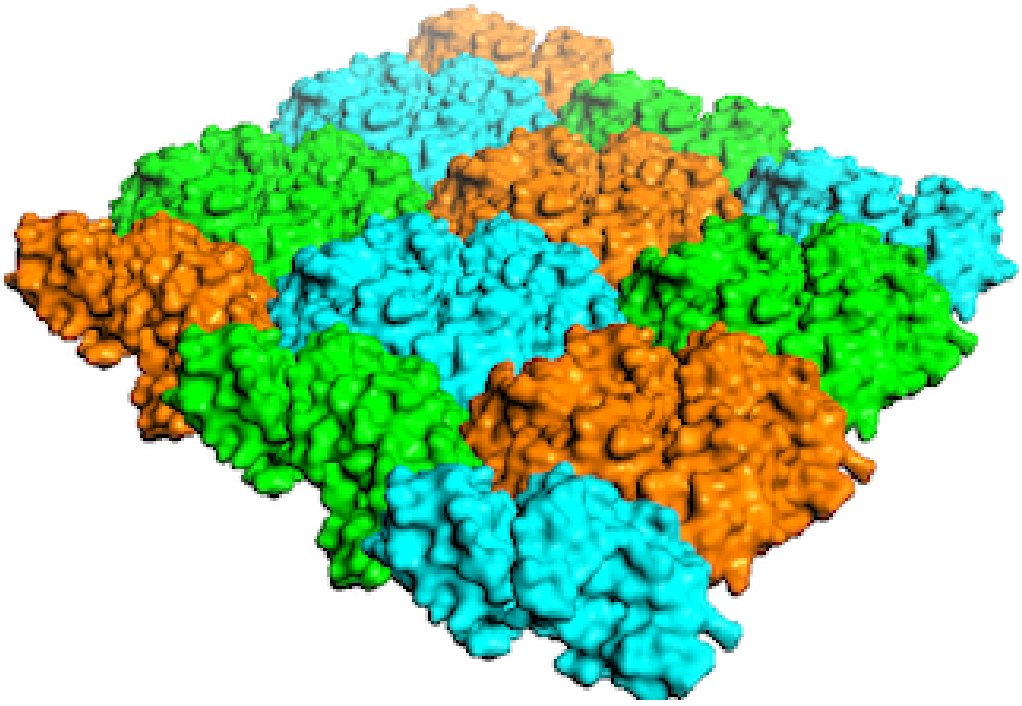

FIGURE 3.2: Surface of a sheet of PduA hexamers, generated from PDB: 3NGK. Showing the concave, or N and C- terminal surface, which is widely thought to face towards the BMC lumen. Individual hexamers are coloured orange, green and cyan to ease differentiation. Each hexamer can be seen to interact with six other hexamers in a symmetrically identical fashion, thereby maintaining six- fold symmetry even when assembling into larger objects.

3.3.1 PduA forms large oligomers when analysed with Native PAGE.

Native PAGE is a useful tool for analysing the size and oligomeric states of proteins. Unlike SDS- PAGE there is no denaturant and so a protein runs in its native state, with intermolecular interactions intact. This allows for inferences to be made regarding oligomeric states and the overall size of complexes. One caveat is that due to the lack of SDS providing a uniform negative charge, the pI of a protein can radically effect its migration. This results in proteins of similar sizes but different pIs running to different positions.

Native PAGE was used to compare the non- tessellating mutant, K26D, with PduA and PduA$_{GST}$ (figure: 3.3). PduA$_{GST}$ runs to a similar position as K26D, indicating

that as with K26D this mutant does not tessellate. Whilst the migration of K26D and PduA$_{GST}$ is similar, the latter runs slightly further, likely due to the large 30- residue extension on the C- terminus of PduA$_{GST}$. Consequently, this produces a large difference in the size of a single hexamer when compared to K26D (the additional seven residues per monomer produces a size difference of 42 residues per hexamer). Additionally, the substitution of the positively charged lysine for negatively charged aspartate on all six K26D monomers will shift the pI of K26D to be more acidic. The greater the difference between the pI of the protein and the pH of the running buffer (8.5), the greater the migration.

It is evident that PduA migrates significantly less than either mutant (figure: 3.3). Additionally, PduA tends to smear and a large amount of the protein is retained in the well whereas, no K26D is visibly retained. PduA and both K26D and PduA$_{GST}$ have similar isoelectric points and are approximately the same size. Therefore, either PduA is running as multimers of hexamers, or both mutants are not forming hexamers, but are instead running as monomers. K26D is known from the crystal structure to form hexamers [25], and both K26D and PduA$_{GST}$ elute at a volume consistent with a single hexamer following SEC (figure: 3.4). It can therefore be inferred that PduA is running as multimers of hexamers, and there are a variety of these multimers present. This heterogeneity accounts for the smeared appearance of the band and substantial staining in the well, which corresponds to much larger oligomers unable to enter the gel matrix.

The SEC data (figure: 3.4) appears to be in conflict with Native PAGE data, as all PduA variants elute at the same point. However, a large amount of PduA is eluted in the void volume of the column, when compared to K26D and PduA$_{GST}$. This indicates that large oligomers are present, beyond the resolution limit of the column. Additionally, substantially less time was required to concentrate both K26D and PduA$_{GST}$ when compared to PduA, even when factoring in concentration and volume disparities. This indicates that PduA is present as larger molecules in the concentrator, which increases the retention time within the reservoir.

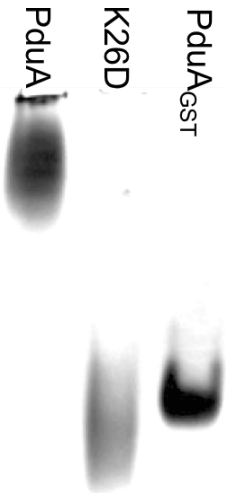

FIGURE 3.3: 4-20% gradient Native PAGE of PduA and the two non-tessellating mutants (K26D and PduA$_{GST}$), both non- tessellating mutants migrate significantly further than PduA . All proteins have similar pIs, it can therefore be inferred that PduA is assembling into a larger oligomer than either non- tessellating mutant. In addition PduA appears quite smeary, indicating the sample is less homogenous than either of the non- tessellating mutants.

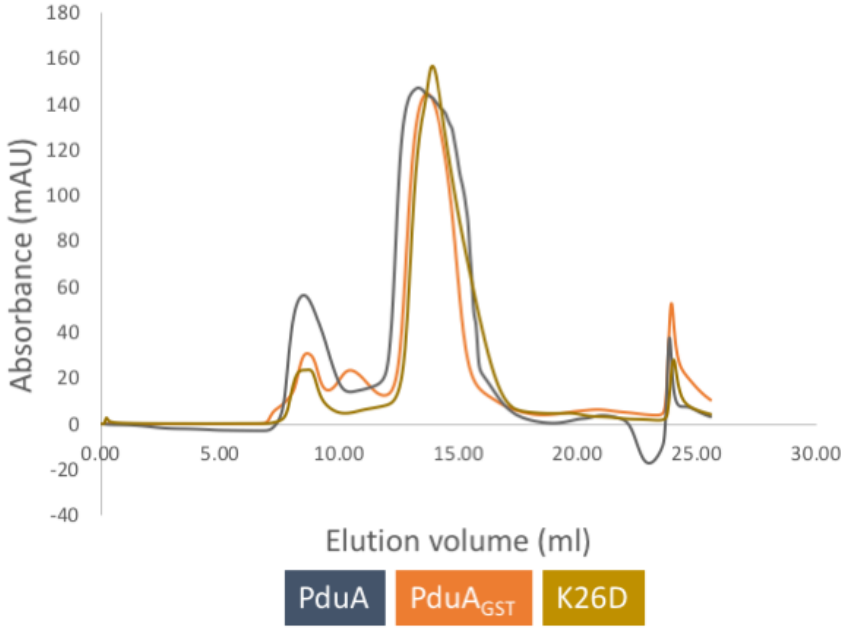

FIGURE 3.4: Comparison of the elution profiles of PduA mutants. All three (PduA, PduA$_{GST}$ and K26D) elute at the same volume, consistent with a hexamer. However, the amount of protein eluted in the void volume (8 ml) is significantly higher for PduA when compared to the non- tessellating mutants. Indicating that, large structures, beyond the resolution limits of the column are present.

3.4 Trapping PduA oligomers using glutaraldehyde cross- linking

In order to determine the precise number of PduA hexamers present in the band seen in figure: 3.3, PduA was treated with glutaraldehyde in order to covalently cross-link the PduA hexamers. Glutaraldehyde features two active carboxyl groups at either end of a five- carbon chain. These carboxyl groups react with primary amines through a condensation reaction, cross linking the two groups [86]. PduA is a particularly good candidate for glutaraldehyde cross linking due to the proximity of the lysine at position 26, which is involved in tessellation (figure: 1.11). Glutaraldehyde should be able to cross- link these lysines, trapping multimers of hexamers, which can then be isolated and analysed. PduA was treated with 0.1 % glutaraldehyde (section: 2.5.1) over a timeframe of 25 minutes, with samples taken at intermediate points. The reaction was terminated with the addition of 100 mM Tris (pH 8.0). To assess the degree of cross- linking over time, these samples were run on SDS- PAGE (figure: 3.5) where the change in different band intensity over time was visualised.

The prominent species present is a ~90 kDa band which corresponds to a single PduA hexamer. Over a prolonged period of exposure to glutaraldehyde the intensity of this band appears to decrease, and increasingly intense staining can be observed observed in the well. This staining is produced by cross- linked oligomers, too large to enter the gel matrix. In addition to this, a variety of other minor bands are present, notably several below 90 kDa. These bands are likely produced through cross linking of several PduA monomers, but not complete hexamers. Two lysines present at positions 12 and 72 are suitably proximal to facilitate this monomer- monomer interaction. Several bands corresponding to molecular weights higher than 90 kDa are also present, however the resolution of these bands decreases at higher molecular weights.

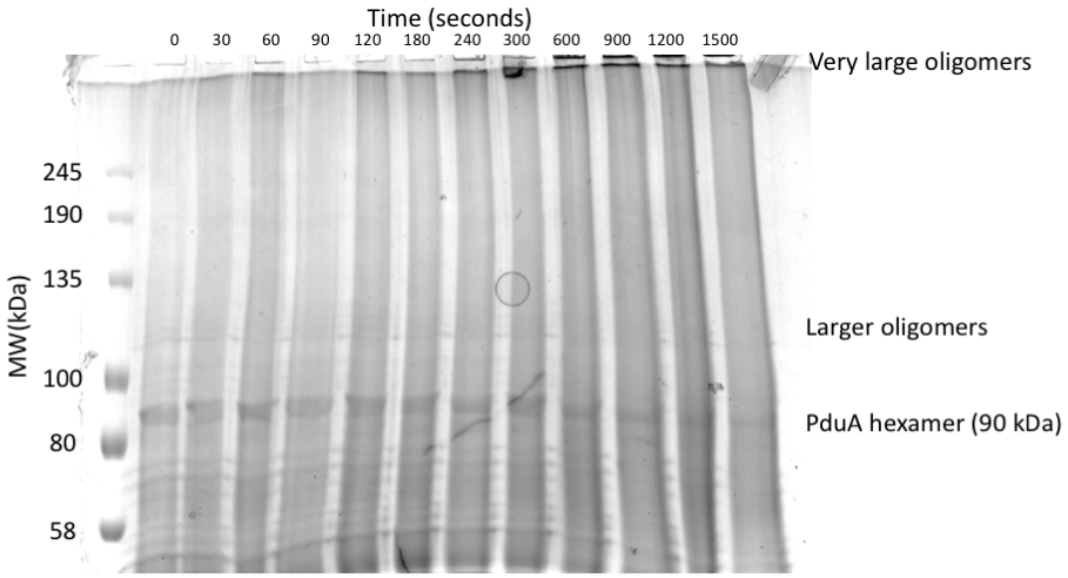

FIGURE 3.5: 8 % SDS- PAGE of glutaraldehyde treated PduA over a timeframe of 1500 seconds. In all lanes, a prominent band can be seen at 90 kDa, consistent with the size of a single PduA hexamer. This indicates the glutaraldehyde has been successful in covalently linking the PduA monomers into a single hexamer (as where non-treated PduA is analysed with SDS- PAGE only a single band at 14 kDa is visible (figure: 3.1)). The loss of intensity of this band over time, and an increase in the intensity of protein retained in the well, suggests that these hexamers are cross- linked further over time into larger assemblies.

3.4.1 MALDI- TOF for identification of cross- linked species

In order to better analyse the larger species present following glutaraldehyde cross-linking of PduA, a sample was taken from the 1500 second sample show in in figure: 3.5 and assessed with MALDI- TOF (figure: 3.6). This technique relies on co-crystallising the analyte (in this case PduA) with a matrix (sinapic acid). The matrix is protonated by a laser and this proton is transferred to the analyte, producing a charged species. The time of flight of this charged species through an electric field can be measured and the mass deduced. One issue with MALDI- TOF (and other Mass spectrometry methods which rely on time of flight) is that they are less suitable for large molecules, due to their being less able to fly and therefore difficult to measure in any meaningful quantity.

In figure 3.6 a variety of masses are detected relating to 1- 12 PduA monomers in the range sampled. This demonstrates the incomplete cross- linking of PduA. Whereby, instead of trapping whole cross- linked hexamers (and multimers of these), smaller multimers are detected, as well as hexamers with several monomers covalently attached. Although these species are detected, it is unlikely they are actually present in solution. This is a result of the denaturing methodologies used, which ensure detection of the smallest cross- linked species, as any non- covalent bonds are disrupted. In contrast, under non- denaturing conditions the interface between PduA monomers is relatively strong and single monomers would not persist for long before oligomerisation into hexamers. Therefore, the only detectable species using non- denaturing conditions should be multimers of whole hexamers.

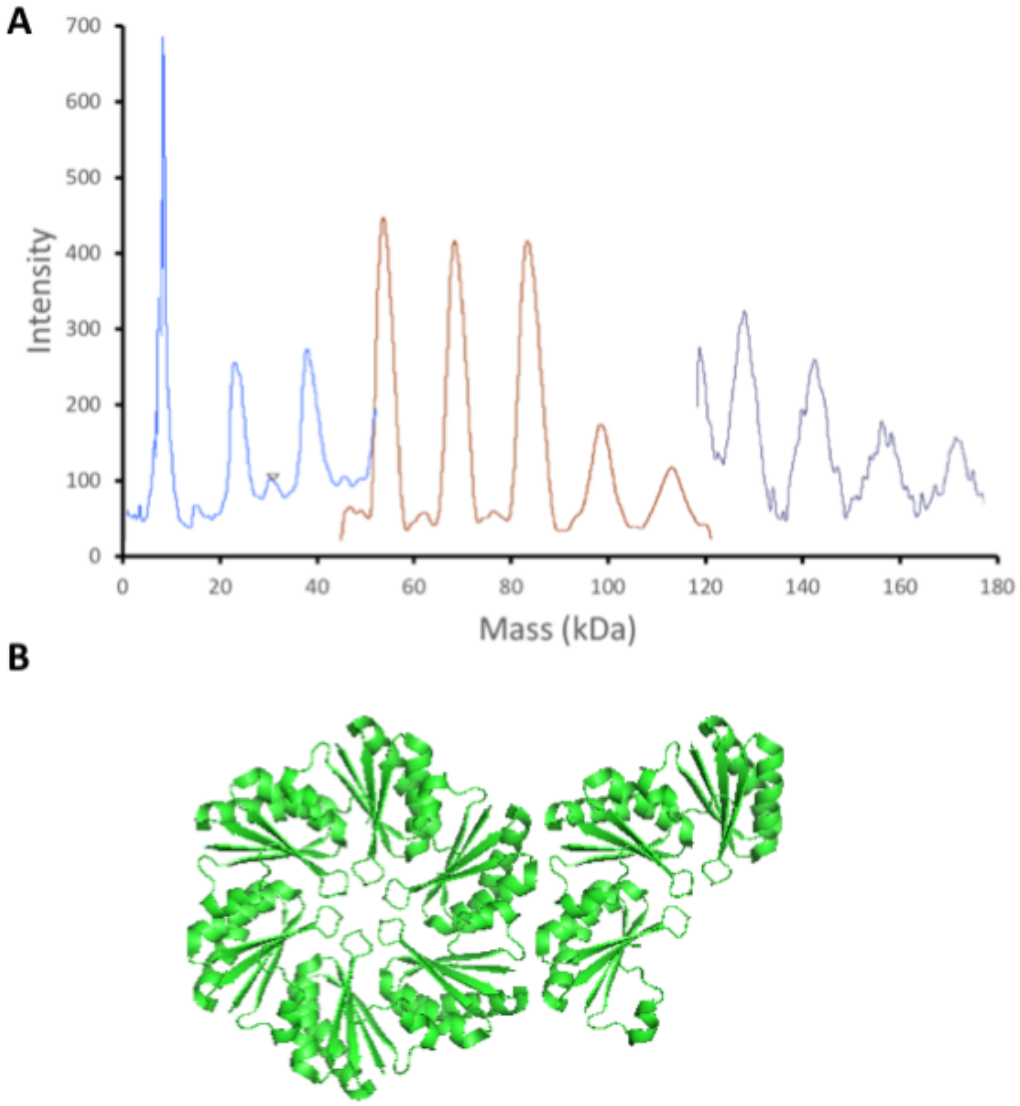

FIGURE 3.6: Mass spectra of PduA treated with glutaraldehyde for 1500 seconds (A). Peaks corresponding to the masses of 1- 12 PduA monomers are visible, confirming the results seen following SDS-PAGE of glutaraldehyde treated PduA. Cross- linked PduA hexamers are present at peaks 90 and 180 kDa, corresponding to one and two hexamers respectively. In addition, partially cross- linked species are present at peaks between 90 and 180 kDa. Here, not all six PduA monomers in a hexamer have been cross- linked into a covalently linked hexamer. Therefore, whilst some inter- hexamer covalent interactions are maintained some intra- hexamers interactions are not, which produces species containing a partial number of whole hexamers. A theoretical example of such a species is shown in panel (B). This species would have a detectable mass of 135 kDa. One hexamer has been totally cross- linked by glutaraldehyde and therefore is detectable as such, however in the other features only three cross- linked monomers and only the mass of these three would be detected.

3.4.2 Native techniques for identification of cross- linked species

To assess whether whole, cross- linked hexamers had been produced, both size exclusion chromatography and Native PAGE were employed. Both of these methods maintain the native state of the protein and therefore, only hexamers or multimers of complete hexamers should be detected.

The SEC profile of cross- linked PduA (figure: 3.7) shows several clear peaks, indicating there are multiple well defined, homogenous species present. The wide variety of masses detected by MALDI- TOF (figure: 3.6) are not visible. The peak at ~8 ml is the void volume of the column and a variety of large oligomers beyond the resolution range of the column are eluted here. Large, discrete oligomers are also present as can be seen from the peak at ~12 ml. This peak also features a shoulder at ~11 ml which may be several slightly larger oligomers that can not be individually resolved. The peak at ~14 ml is the same as that seen for untreated PduA (figure: 3.4), indicating that some PduA remained as a single hexamer.

Analysis of the SEC fractions eluted at 11, 12, 13 and 14 ml from SEC (figure: 3.7) using Native PAGE (figure: 3.8) shows four bands. The top band in the lane labelled 11 appears more well defined. However, this is the bottom of the well and corresponds to multiple large oligomers which could not enter the gel matrix.

The elution volume of untreated PduA is typically ~14 ml, and there was a large peak at this volume in figure: 3.7. When this fraction is run on Native PAGE (figure: 3.8), the major band is towards the bottom of the gel. Therefore, it is likely that this band represents a single PduA hexamer, as both elute at the same volume. The simplest explanation as to the identity of the other bands and peaks are that they correspond to two and three cross- linked PduA hexamers.

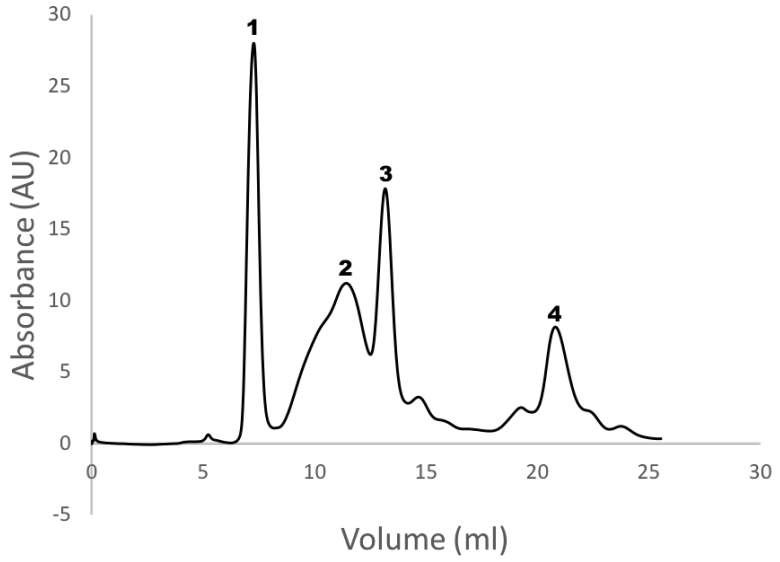

FIGURE 3.7: S200 elution profile of PduA treated with 0.5 % glutaraldehyde for 10 minutes, 4 peaks are clearly visible, corresponding to several different PduA oligomers with different masses. A slight shoulder is prominent on the peak labelled 2, indicating that this peak may contain several different oligomers of similar sizes. The peak labelled 1 is the void volume of the column, and so despite appearing quite sharp it is likely there are several different oligomers present beyond the resolution of the column. The peak labelled 4 elutes at 22ml and is likely some contaminant, as it is smaller than a PduA monomer (as determined by the calibration, section: 2.4.4).

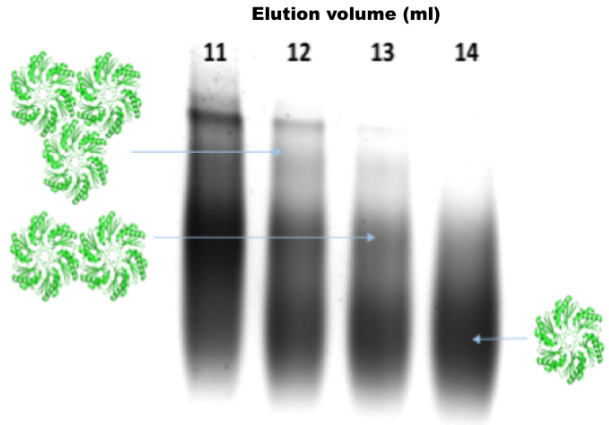

FIGURE 3.8: Native PAGE of SEC fractions presented in figure 3.7. The elution volume from an S200 column is shown above the gel. The migration of the bands increases the later the fraction was eluted following SEC. As was apparent with the SEC data, there appears to be at least three species present.

3.4.3 Verification of the hypothesised trimer of hexamers

To verify that the proposed monomer, dimer and trimer of PduA hexamers presented in figure: 3.8 are the species responsible for the three bands observed following SEC and Native PAGE of cross- linked PduA, MALDI- TOF was employed again. On this occasion instead of scanning wide regions of masses, only the regions which correspond to the masses of one, two and three PduA hexamers were probed, in order to determine if these species were present in the sample.

Here (figure: 3.9), peaks clearly relating to the masses of one, two and three hexamers are present. There is no peak corresponding to four hexamers. However, one limit of MALDI- TOF is that larger proteins are less able to fly and thus are less likely to be detected. This also likely relates to the high intensity seen for the single hexamer and does not necessarily indicate their being present in greater abundance. These spectra confirm that the monomers, dimers and trimers postulated to be formed following glutaraldehyde cross- linking are indeed present. How these trapped assembly intermediates relate to the variation observed in the sizes of PduA and its non- tessellating mutants was assessed using Native PAGE (figure: 3.10).

Native PAGE (figure: 3.10) revealed that, the PduA_{GST} mutant shows similar migration when compared to the lowest band present following PduA cross-linking. This band has been identified as a single hexamer, confirming that PduA_{GST} runs as a single hexamer. The migration of PduA is more difficult to interpret, as the smearing produced by the formation of larger complexes inhibits the identification of a defined PduA band. A region of this smear towards the bottom is more intense. This is presumably the more abundant PduA multimer and migrates to similar position as the top cross- linked PduA band, which has been identified as a trimer, or possibly tetramer of PduA hexamers.

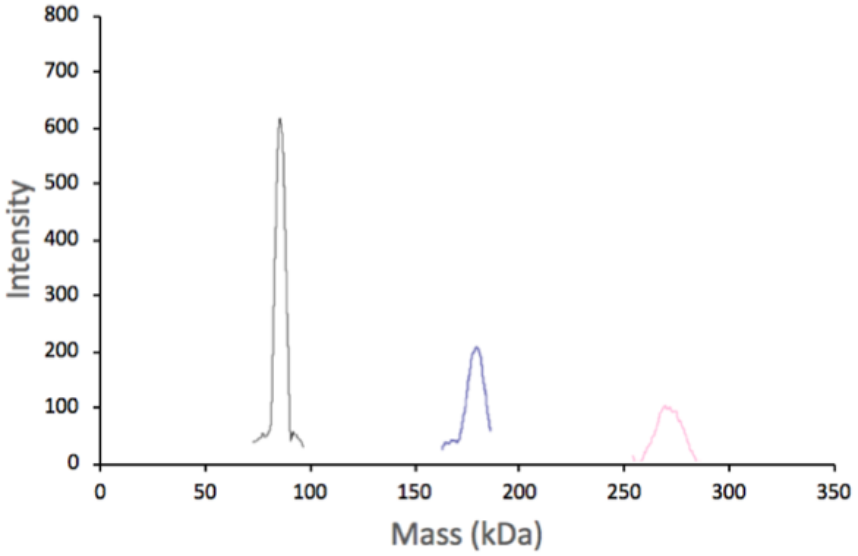

FIGURE 3.9: Mass spectra of cross- linked PduA. Specifically searching for peaks in the 70- 100, 150- 175, 225- 275 and 300- 350 kDa ranges. Peaks in this range were present, indicating that cross- linked dimers and trimers of PduA hexamers are produced following treatment of PduA with glutaraldehyde.

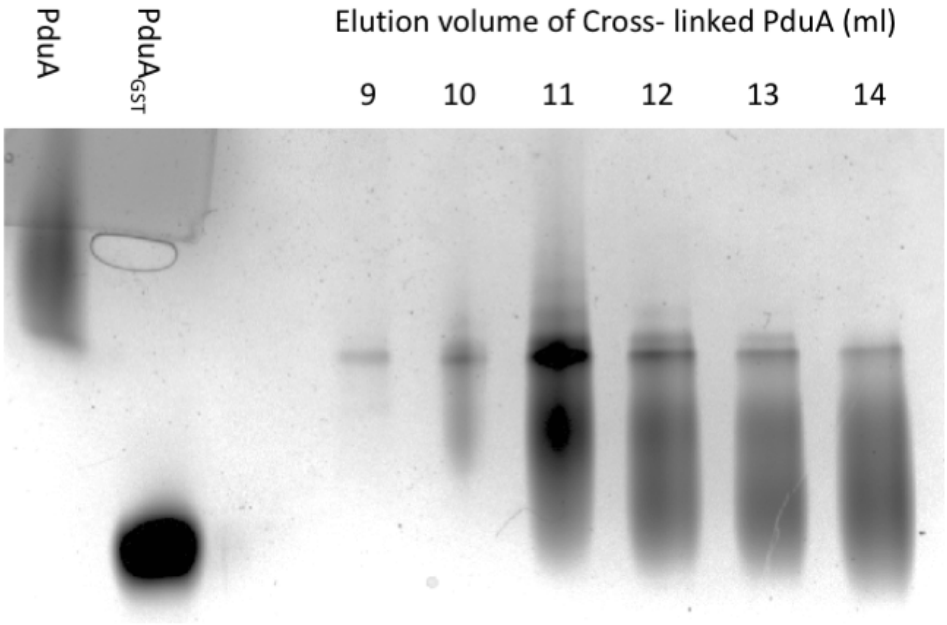

FIGURE 3.10: 4- 15% gradient Native PAGE of PduA, PduA$_{GST}$ and the cross- linked eluted following SEC (shown in figure 3.7). PduA$_{GST}$ shows similar migration to lowest band, which was eluted at a mass consistent with a single hexamer. This corroborates the idea that the smallest detectable species of cross- linked PduA is a single hexamer. PduA migrates similarly to the uppermost band of cross-linked PduA, which is speculated to be a trimer of hexamers.

3.5 Trapping PduA oligomers using mutagenesis

Due to the PduA- PduA hexamer interface interaction being mediated by lysine 26 [25], the oligomerisation of PduA hexamers could potentially be manipulated by controlling the presence or absence of this residue in a single PduA hexamer. In order for this to be feasible however, a single chain concatenated PduA (with each monomer joined at the N and C- termini) must first be created. Due to the entire PduA hexamer being encoded on a single gene this allows for the alteration of certain residues in one PduA monomer and not in the others.

Such a mutant was created by Dr Ruth Rose, in the form of $PduA_6$, where six PduA monomers are linked by a (Gly- Gly- Ser- Gly- Thr) x6 linker. This 36 residue linker facilitates the oligomerisation of linked PduA monomers into a single PduA "pseudohexamer". The trimeric (PduB, PduT) shell proteins feature similarly concatenated BMC domains, but with two BMC domains per polypeptide chain rather than six.

A $PduA_6$ variant, $PduA(K26D)_4(WT)_2$, referred to as $K26D_4$, was created. $K26D_4$ features lysine 26 on only two adjacent PduA "monomers", whilst all other lysine 26 residues are mutated to aspartate (which has been previously demonstrated to inhibit tessellation [25]). It is predicted that this mutant should only form interfaces between the unaltered lysine 26 residues, trapping the PduA trimer of hexmers following assembly. This trimer can then be compared to the PduA oligomers that occur naturally and have observed using Native PAGE (figure: 3.3). A diagram of the oligomer intended to form is shown in figure: 3.11.

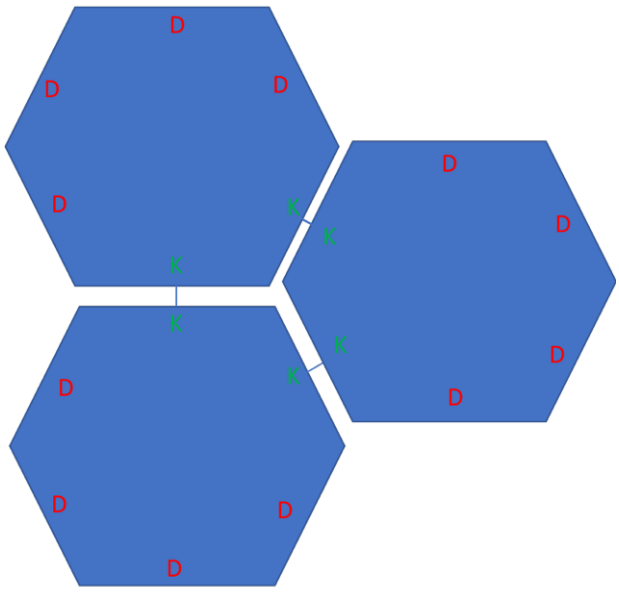

FIGURE 3.11: Diagram of the trimer of PduA hexamers anticipated to be formed by K26D$_4$, the interactions between antiparallel lysines results in aspartate being presented on the outer edges of the trimer, which is not conducive to further oligomerisation.

3.5.1 Analysis of concatenated PduA mutants

Comparisons between the concatenated K26D$_4$ mutant and PduA can be used to confirm the size of the oligomers formed by PduA (figure: 3.8). As K26D$_4$ should not be able to tessellate beyond a trimer of hexamers, this mutant should migrate similarly to PduA on a Native PAGE, if PduA does tessellate into trimers of hexamers.

Initially, SEC was performed on K26D$_4$ to assess whether the polypeptide chain was assembling into a pseudohexamer (figure: 3.12). The SEC elution profile was consistent with that of PduA. Indicating that, the polypeptide chain was assembling into the anticipated flat pseudohexamer, comparable in overall structure to PduA. SDS- PAGE of the purified concatenated mutants (figure: 3.13A) yielded a prominent band at ~75 kDa, representing the entire polypeptide chain. However, several minor bands can also be seen down to ~12 kDa (single monomer), indicating the degradation of the concatenated variants was occurring. Despite this, the cleavage

of the linking region should not be too problematic due to the high stability of interactions between monomers within the hexamer, which maintain the overall tertiary structure.

Native PAGE of Both $PduA_6$ and $PduA_{GST}$ are shown in figure: 3.13B. As expected $PduA_6$ shows similar migration to $PduA_{GST}$. This indicated that the concatenation into a single chain did not cause the protein to tessellate. There was notably more $PduA_6$ retained in the well, indicating an increased propensity for the concatenated variant to aggregate.

$K26D_4$ was anticipated to show similar migration to the major PduA band (figure: 3.13C), albeit more well defined (as it should not be able to assemble into oligomers greater than three). However, $K26D_4$ showed a strange pattern of migration. Initially, two bands were visible, one at the same position as K26D and one at a slightly lower position than PduA. This is to be expected if they are the same mass, as the substitution of four positive arginines for negative aspartates will result in an increase in the net negative charge by four per pseudohexamer, producing an increase in migration. A similar phenomena is observable with the monomeric K26D and $PduA_{GST}$ variants where, despite being the same mass, the difference in charge results in K26D migrating marginally further. This effect was shown earlier in this chapter (figure: 3.3). Native PAGE analysis of the same sample over time (figure: 3.13C) resulted in the top band (hypothesised to be a trimer of hexamers) increasing in intensity. Conversely, the intensity of the bottom (single hexamer) band decreased. It is therefore likely the mutant formed the anticipated trimer, but at a decreased rate when compared to PduA. What is more surprising, is the appearance of a poorly defined band above that corresponding to the trimer. This is often seen in older samples of PduA and is thought to contain larger oligomers and the beginning of tubes or sheets. The high molecular weight, combined with high heterogeneity between oligomers results in the streaking and a poorly defined band. The availability of only two lysines per pseudohexamer in the $K26D_4$ mutant should restrict the formation of oligomers larger than three. However, the presence of protein oligomers larger than a trimer of hexamers is indicated by the streaking observed.

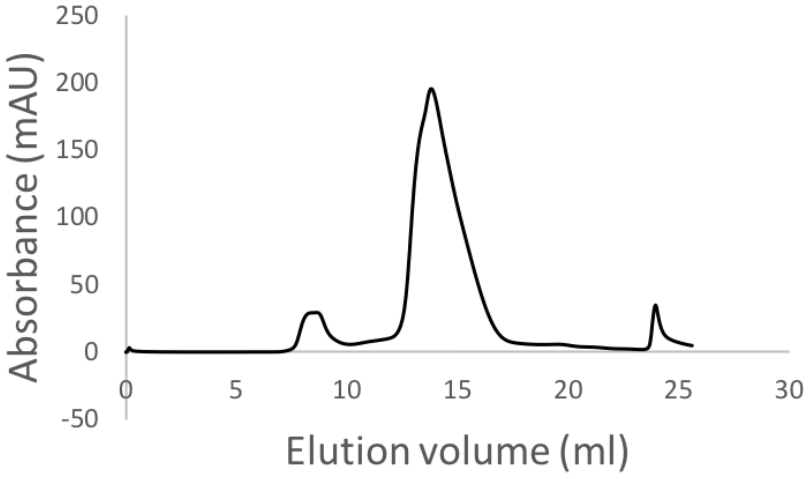

FIGURE 3.12: Size exclusion chromatography of $K26D_4$. The elution volume (~14 ml) is consistent with the earlier reported elution volume of PduA (figure: 3.4). Indicating that, $K26D_4$ does assemble into the anticipated pseudohexamers.

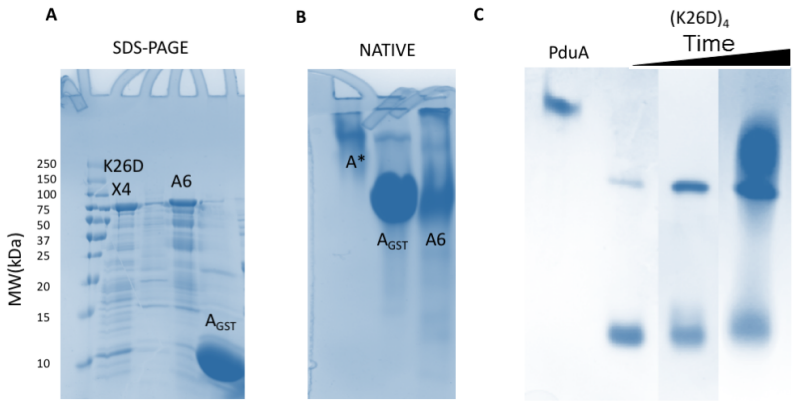

FIGURE 3.13: SDS- PAGE of $PduA_{GST}$, $PduA_6$ and $K26D_4$(A). The presence of a band at 75 kDa demonstrates the concatenation of both $PduA_6$ and $K26D_4$ was successful. There are numerous bands present below this 75 kDa band however, suggesting that the single chain PduA constructs are unstable and prone to degradation. A 4-15% Native PAGE of PduA, $PduA_{GST}$ and $PduA_6$(B) shows that, $PduA_{GST}$ and$PduA_6$ migrate to the same position and concatenation does not influence the propensity to tessellate. A 4-15% Native PAGE of PduA and $K26D_4$ (C) shows that surprisingly K26D does not migrate to the same position as PduA as anticipated, but instead is present as two distinct bands, relating to two distinct oligomers. Analysis of the same sample over time results in the intensity of the top band increasing as is shown in panel C where three different gels taken 2 weeks apart are compiled.

3.5.2 Accounting for the tessellation seen in the K26D$_4$ mutant

Assembly of K26D$_4$ into oligomers larger than three pseudohexamers (figure: 3.13C) is unexpected. The most plausible explanation is that lysine 26 does not exclusively interact with opposing lysine 26. Previous work demonstrated that K26D mutant did not tesselate due to two primary reasons [25]:

1)The anti-parallel lysine interaction is one of the driving forces of PduA assembly.

2) The repulsive nature of two proximal negative (aspartate) residues is not be conducive to an interaction.

It is this second feature of the K26D mutant which was overlooked. If the K26D mutant is overlaid on tessellating PduA hexamers (figure: 3.14), the position of the aspartate with respect to the lysine is conducive to the interaction of WT PduA and the K26D mutant. The NZ atom of lysine 26 is 2.6 Å from the carbonyl oxygen atom of aspartate 26 and 3.9 Å from the OD1 atom. This proximity may allow for hydrogen bonding to occur and perhaps even a salt bridge between the positive amino group of lysine and negative carboxyl group of aspartate. In contrast the distance between the amino and carboxyl groups of the anti- parallel lysines in native PduA is 2.8 Å. Therefore, the K26: D26 interaction is very likely permissible.

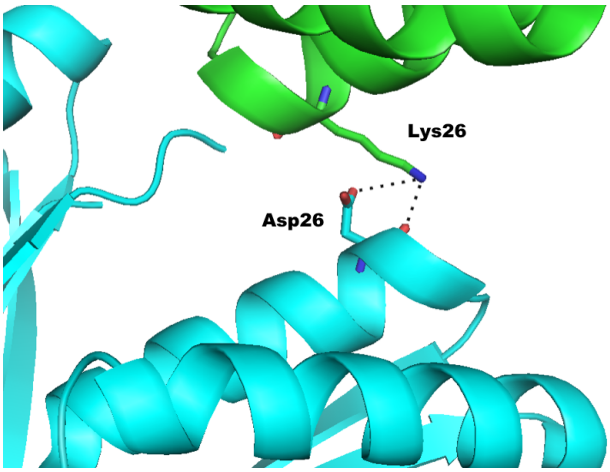

FIGURE 3.14: Cartoon representation of the crystal structure of K26D (cyan) overlaid onto a sheet off tessellating PduA (green). The proximity of lysine 26 and aspartate 26 would allow for tessellation of the mismatched edges of WT PduA and K26D to be permissible.

Given this revelation, the potential oligomers which the $K26D_4$ mutant can form are easily predictable as there are two conditions which must be satisfied:

1) Lysine can feasibly pair with either another lysine or aspartate.

2) Aspartate cannot pair with another aspartate due to the repulsive electrostatic interaction.

A diagram of the largest possible oligomer formable by $K26D_4$ is shown below (figure: 3.15).

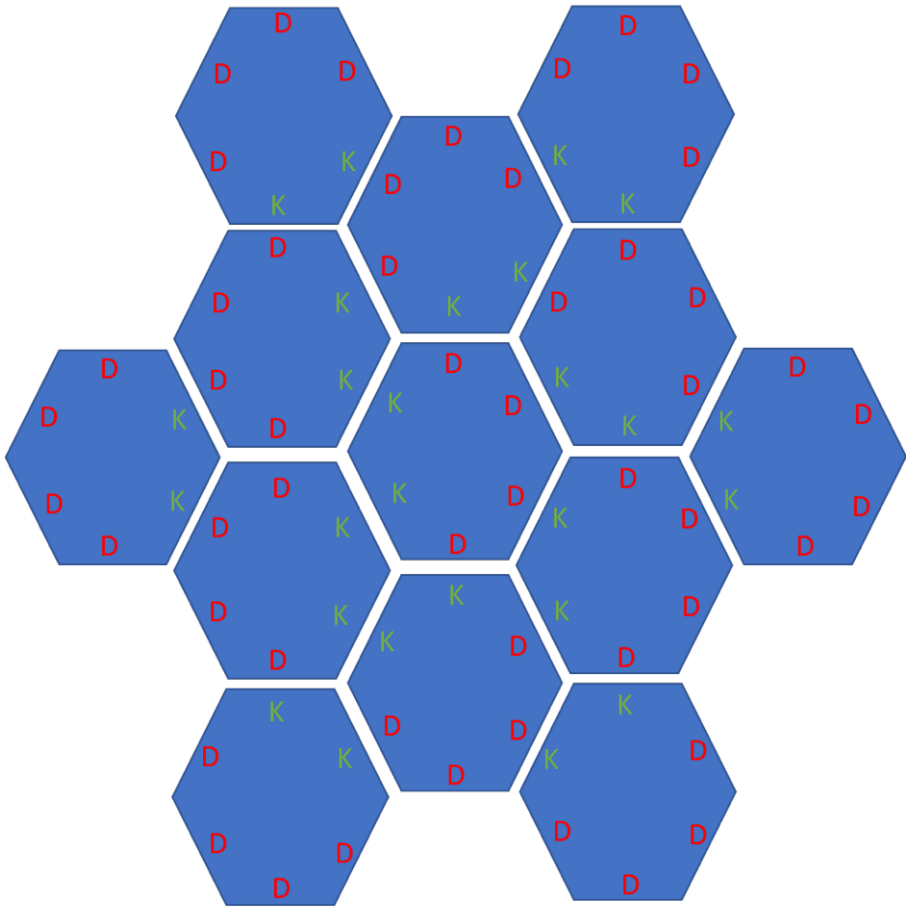

FIGURE 3.15: The largest possible oligomer likely formable by $K26D_4$ assuming lysine 26 can interact with asparate 26. Here, lysine 26 can pair with either another lysine or Aspartate. However, aspartate 26 may only pair with Lysine. Oligomers larger than this would be possible if the lysine of a single hexamer were to bind to one of the outer Aspartates, however only a single edge could be satisfied.

3.6 Investigating macrostructures formed by PduK

PduK is a hexamer, featuring a high degree of amino acid sequence homology to PduA (figure: 3.16). Several shell proteins, including PduA have been shown to arrange into macrostructures, in the form of tubes or sheets. Given this high degree of sequence homology and the conservation of the lysine 26 residue which facilitates tessellation of PduA, the formation of macrostructures by PduK was investigated. In addition to the BMC domain (which it shares with PduA), PduK features a large ~50 residue C- terminal extension (figure: 3.16). This extension serves no known purpose, although several BMC shell proteins feature them. In light of this a truncated variant of PduK (A100>STOP) was produced, which featured only the BMC domain. The effect of the presence or absence of this C- terminal domain on macrostructures formed by PduK is investigated in this section.

```
PduK -MKQSLGLLEVSGLALAISCADVMAKAASITLVGLEKTNGSGWTVIKIIGDVTSVQAAIS 59
PduA MQQEALGMVETKGLTAAIEAADAMVKSANVMLVGYEKI-GSGLVTVIVRGDVGAVKAATD 59

PduK TGVSFADQRDGLVAHKVISRPGDGILSRSVPAESEQTPAPTPVVPHEEMTGDHAIPEAPQ 119
PduA AGAAAARNVGEVKAVHVIPRPHTDVEKI-LPKGIS------------------------- 93

PduK DAGLISCNLCLDPACPRQKGESRSLCLHSGKRGEA 154
PduA ------------------------------------
```

FIGURE 3.16: Amino acid alignment of PduA and PduK. The thee residues which have been implicated in the tessellation of PduA: lysine 26, valine 51 and arginine 79 are conserved across both proteins (highlighted in green). PduK features a large 50 residue extension, which is not homologous to any known protein domain (highlighted in red).

3.6.1 Purification of PduK

Purification of PduK proceeded in the same manner as that for other shell proteins, outlined in chapter: 2, section: 2.3.5. Following Nickel IMAC, SDS- PAGE analysis of the eluted fraction suggested a band at ~26 kDa, surprising as PduK has a theoretical mass of ~16 kDa. Mass spectrometry was performed in order to confirm the mass of the eluted species and ensure it was PduK and not some other contaminant. Following the detection of a peak at ~16 kDa confirming the presence of PduK, SEC

was performed. Here, a small peak eluted in the void volume of the column suggesting a small degree of aggregation. Additionally, a single major peak was present at 14.5 ml, consistent with the mass of a hexamer calculated from earlier equilibration of the column. Following SEC, fractions containing pure protein were pooled and concentrated before further SDS- PAGE analysis (figure: 3.17), revealing a pure, highly homogenous sample of protein.

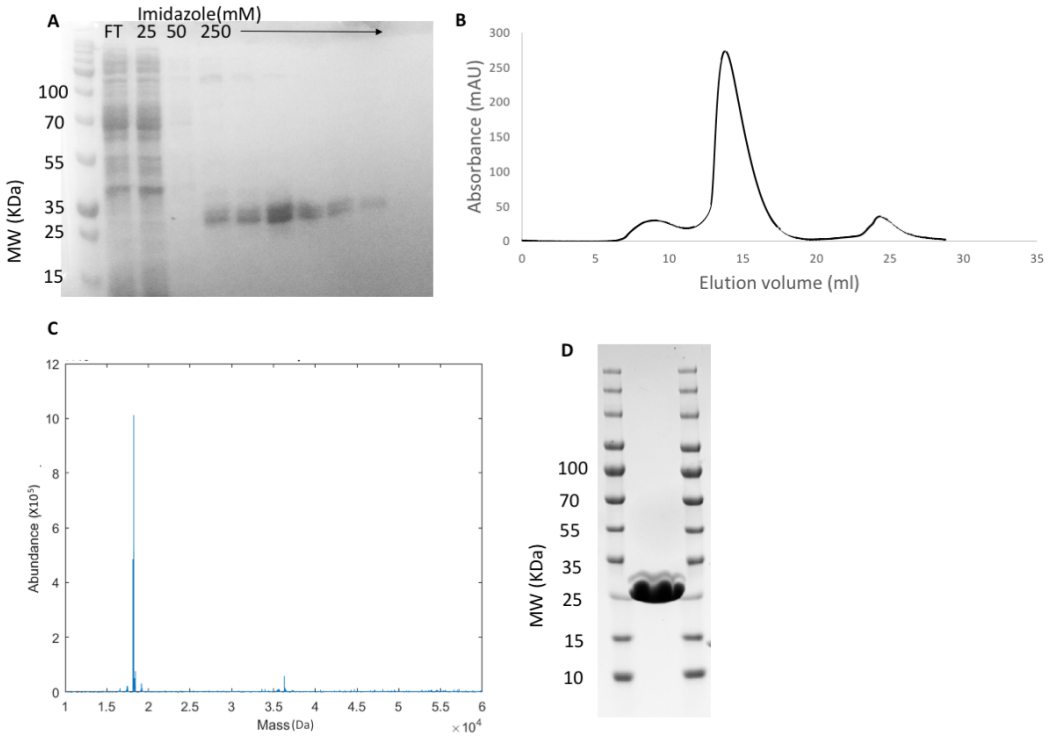

FIGURE 3.17: Purification of PduK, showing SDS- PAGE of the IMAC fractions (A) PduK elutes at 500 mM imidazole at a mass of ~26 kDa. This deviates significantly from the theoretical mass of 16 kDa. SEC using an S200 column (B) yields an elution profile with major peak at 14.5 ml, corresponding to a mass consistent with a single hexamer. Mass spec (C) showed that the purified sample was consistent with the theoretical mass of PduK (16 kDa). Therefore, the mass seen following SDS- PAGE must be some artefact of PduK migration through the acrylamide matrix. Finally SDS- PAGE of the fraction eluted from SEC (D) shows protein was purified to high homogeneity.

3.6.2 Production of truncated PduK

In order to investigate the role of the PduK C- terminus, a mutant featuring only the conserved BMC domain, with residues 100- 152 removed was produced (A100>STOP). Following site directed mutagenesis, and confirmation of the desired mutation though sequencing, it was transformed into BL21(DE3) cells and protein was expressed and purified. Analysis with SDS-PAGE demonstrated that the truncated protein was being produced in reasonable quantities, to a high homogeneity (figure:

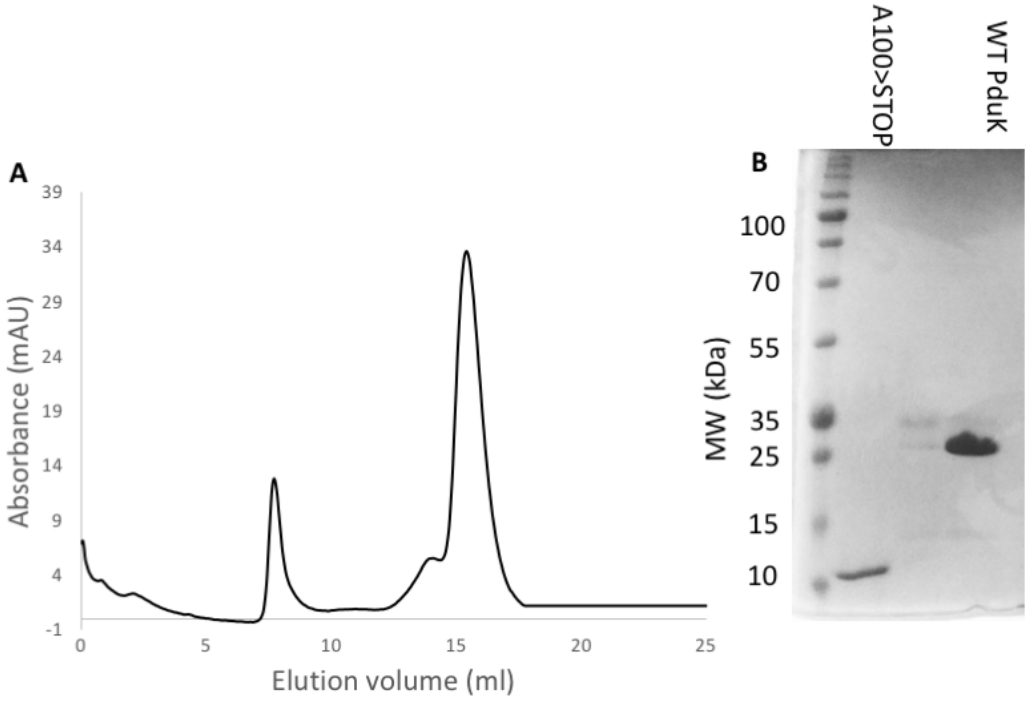

FIGURE 3.18: Purification of PduK A100>STOP. SEC using an S200 column (A) yields an elution profile with major peak at 15.5ml, slightly later than full length PduK, due to the shorter sequence length, but still indicative that the truncated mutant assembles into a single hexamer. SDS- PAGE of A100>STOP and WT PduK (B), shows that the mutagenesis was successful and that the truncated variant runs to approximately 12 kDa following SDS- PAGE.

3.6.3 The C- terminus of PduK effects the formation of higher order structures when assessed with TEM.

Following the production of PduK and PduK A100>STOP both were concentrated to 5 mg/ml and visualised using TEM. It is widely known that several microcompartment shell proteins assemble into tubes and/ or sheets at high concentrations, and so analysis with TEM should demonstrate if the absence of the C- terminus has any effect on formation of higher order structures.

PduK was observed to form large sheets with a crumpled morphology under TEM (figure: 3.19A). At various points, divisions between these sheets were apparent, where interfaces between adjacent hexamers had broken, resulting in some disassembly. These structures appear incredibly uniform, with no clearly identifiable individual hexamers apparent. Many of these structures were over ~1- 1.5 μm across. The diameter of a single hexamer is 6nm and so these sheets would be composed of almost 300 hexamers in any one direction. There are crumpled regions present on the surface visible. These may either be areas where multiple sheets are arranged over one another, or possibly areas where angles have been induced at the hexamer-hexamer interface. This would produce sharp curves in the overall geometry of the structure.

TEM was then repeated using 5 mg/ml A100>STOP. The difference between the structures formed where the C- terminus of PduK is removed is remarkable (figure:3.19B). The tubes formed (circled in red) are reminiscent of those formed by PduA and PduB, albeit with more diverse and slightly larger diameters. What is most interesting is the compartment like structures observed. These are strikingly similar to those formed in vivo where all shell proteins are expressed together [18]. Additionally, the diameter of these compartments is also very similar, at ~80 nm. As seen from the images these objects are incredibly heterogenous, appearing to have two distinct morphologies of tubes and pseudo- compartments. The reason for this is unclear as the BMC domain of PduK shares a high degree of similarity to that of PduA (80 %). Additionally, all residues known to be implicated in tessellation (Lys 26, Val 51 and Arg 79) are conserved. Therefore, one would assume PduA and PduK

A100>STOP would behave similarly.

The differences between the morphologies of full length and truncated PduK are also difficult to rationalise. One possibility is that presence of the large C- terminal extension on all six monomers per hexamer causes a degree of stearic hinderance. This may be sufficient to prevent the curvature between hexamers that would be required to facilitate the formation of objects seen in figure:3.19B. The issue with this explanation is that it would require the C- terminus of PduK to be on the interior of the objects. This was disputed in section: 1.6 due to the recent mounting evidence for the C- termini of hexamers being present on the exterior of the compartment. Another explanation is that the C- terminus of PduK is involved directly in the oligomerisation of hexamers into flat sheets. The mechanism for this is unclear however, if the C- terminus were to obstruct the interfaces between adjacent hexamers, this could potentially limit the curvature of the sheets. This would account for the disparity seen in the curvature of objects formed by full length vs. truncated PduK.

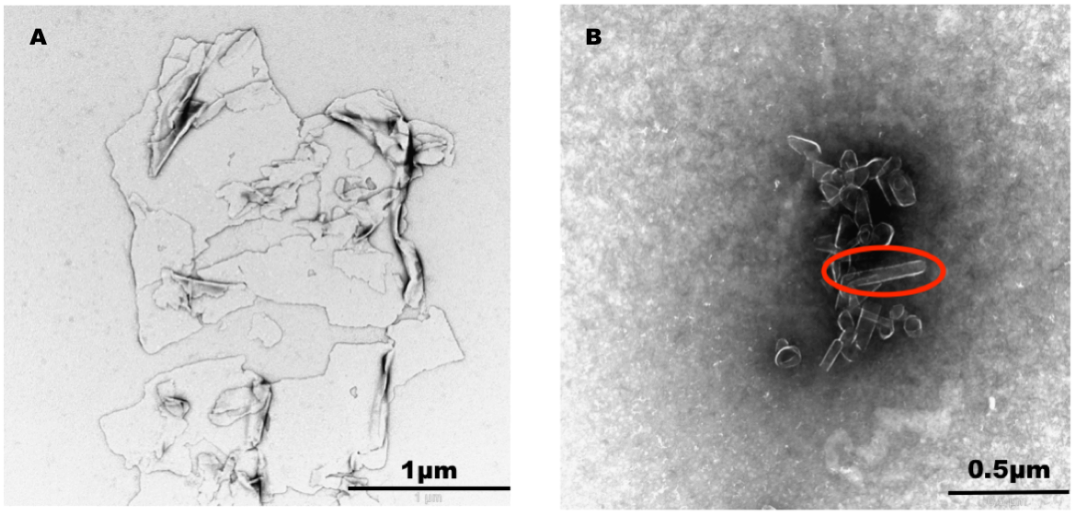

FIGURE 3.19: PduK can be seen to form large flat sheets when visualised using TEM (A). Therefore, as with PduA PduK is able to self assemble. PduK A100>STOP TEM (B). A100>STOP can be seen to form small, highly heterogenous objects when visualised using TEM. The absence of the C- terminal domain results in distinctly different morphologies of the macrostructures formed by PduK.

3.7 Conclusions and Discussion

It has previously been posited that PduA exists as single hexamers in solution which tesselate into sheets and tubes once, either a critical concentration of protein is attained, or the concentration of NaCl is lowered to <0.05 M [25]. In this chapter it has been clearly demonstrated that PduA does not exist as single hexamers in solution before oligomerisation into higher order structures. Instead, smaller assemblies of multimers, likely trimers and tetramers of hexamers are present. These smaller assemblies have been previously undetected due to the methods by which the formation of PduA tubes are monitored. Namely, the appearance of the solution shifting from transparent to white- opaque as insoluble tubes are formed, and TEM to detect and image the large tubes (a technique which would not detect smaller soluble oligomers). Native PAGE is a simple yet effective technique for assessing the self- assembly of PduA, and is capable of detecting multiple different oligomers in a sample. These samples would appear heterogenous, when assessed with other techniques.

Amongst the consequences of PduA existing as oligomers almost immediately following purification, and at high NaCl concentrations are the inferences that can be made from pull down assays. Where it was previously proposed that PduA existed as single hexamers in solution, the logical conclusion of an enzyme co- eluting with PduA following a pull- down assay was that the binding site must exist within a single hexamer. All work done on modelling the PduA- enzyme complex has been completed with this axiom in mind, and so all models to date show an interaction between a single PduA hexamer and enzyme. The work discussed in this chapter has demonstrated the existence of small PduA oligomers. Therefore, the possibility of the enzyme binding site spanning multiple PduA hexamers is plausible, an idea which is expanded on in the following chapters.

The observation that PduK self assembles into sheets is new. More interesting is the radical effect that the removal of the C- terminus has on the morphologies of macrostructures formed by PduK. The precise mechanism underpinning this change is unclear. However, it is interesting that the objects formed by truncated PduK are

similar to those previously reported to be formed by PduA [18, 25, 30]. Therefore, the C- terminus of PduK clearly facilitates the formation of flat, rather than curved sheets and perhaps this plays a role in the formation of the flat facets of the BMC.

3.8 Further work

It is intuitive that, before a large complex can form, the building blocks of that complex must first come together. Therefore, the observation of the assembly of small oligomers of shell proteins is key to understanding how BMCs form as a whole. Here, it has been shown that PduA forms small homo- oligomers of hexamers, prior to forming large tubes and sheets. However, in reality BMCs are hetero- oligomers and it would be therefore be prudent to study the formation of these from dilute heterogenous solutions of different hexamers.

Similar experiments could be undertaken as have been described in this chapter, although as opposed to forming small homologous multimers, several different shell proteins could be added together and analysed. Glutaraldehyde cross- linking experiments may be of significant benefit here, as the conserved interfacial lysines should be suitably proximal to facilitate cross- linking of different shell proteins in a complex. SEC would not be suitable for the separation of these species, due to the high similarity between the overall size of different shell proteins. For example, a trimer of PduA hexamers would be largely indistinguishable from a trimer consisting of two PduA hexamers and a PduB pseudohexamer when assessed using SEC. Different shell proteins do feature a range of isoelectric points and separation based on charge rather than size may be more suitable. Such techniques include, ion- exchange chromatography and isoelectric focussing .

Identifying the relative frequency at which certain hetero- oligomers occur would further illuminate which shell proteins interact more favourably with others. This would allow for the elucidation of how the shell proteins are arranged in the Pdu BMC. Whilst some work has established which shell proteins will interact with others [18], there is currently no information available regarding how preferable these interactions are.

Chapter 4

The crystal structure of PduT from *Clostridium phytofermentans*.

4.1 PduT Introduction.

In the previous chapter the hexameric shell proteins PduA and PduK were discussed. This chapter will focus on the trimeric shell protein, PduT. As with PduA and PduK, PduT features six- fold symmetry and six BMC domains. However, each polypeptide chain of PduT features a tandem repeat of two BMC domains. This tandem repeat allows for the maintenance of six fold symmetry and causes the PduT monomers to assemble as a kind of "pseudohexamer". By having two BMC domains per PduT monomer, this allows for slight variation between the domains, which is not possible in hexameric shell proteins.

Collaborators at the University of Kent supplied data collected from a crystal of PduT from *C. phytofermentans*. PduT from *C. freundii* and *S. typhimirium* have previously been shown to contain a 4Fe- 4S cluster [20, 24]. Although, when the protein is crystallised no such cluster is apparent, likely a product of it being unstable. There is however, a suitable location for a 4Fe- 4S cluster in the pore of the pseudohexamer. Here, three of the four irons are co- ordinated by a cysteine and one iron remains free, this free iron is speculated to bind some other component of the Pdu BMC [20]. Sequence analysis of PduT from *C. phytofermentans* suggests it contains six cysteines in the pore region, rather than the usual three. These additional cysteines were speculated to stabilise the iron- sulphur cluster, allowing it to

be trapped and resolved in the crystal structure. Indeed, the crystals formed were of a brown colour, indicating the presence of an iron sulphur cluster which could potentially be resolved in the crystal structure.

4.2 C. phytofermentans PduT does not tesselate or form sheets in the crystal structure.

Diffraction data was collected for PduT by colleagues from the University of Kent to a resolution of 2.84 Å, at Diamond light source's I04 beamline. The data was supplied as an auto- processed mtz file, processed on site using XIA 3dii. Molecular replacement and refinement were performed using the phenix software suite [87], using the structure of PduT from *S. typhimirium* (PDB:3N79) [24] as a search model. Refinement statistics are shown in table: 4.1. In the final crystal structure there were three copies of PduT per asymmetric unit. Remarkably, unlike the crystal structures of other shell proteins, PduT was not found to form sheets in the crystal lattice. Instead, the trimers were arranged in closely packed pairs, with the Convex faces packing into one another. These pairs were then loosely assembled in the lattice with minimal contacts between one another (figure: 4.1).

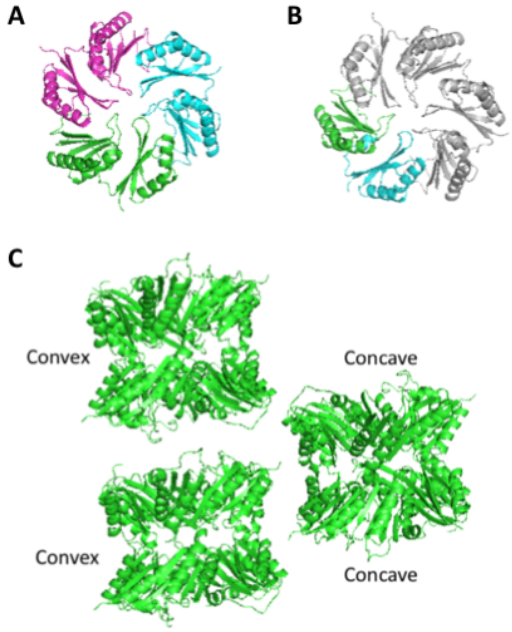

FIGURE 4.1: Crystal structure of PduT, showing of the three each polypeptide chains within the trimer in a different colour (A). Each of these polypeptide chains feature two BMC domains, shown in green and cyan (B). The appearance of PduT in the crystal lattice (C), demonstrates a lack of the typical tightly packed sheets formed by BMC shell proteins

TABLE 4.1: Refinement statistics of the completed PduT structure and crystallographic data.

Wavelength	0.9795
Resolution range	45.52 - 2.75 (2.849 - 2.75)
Space group	P 31 2 1
Unit cell	122.41 122.41 68.09 90 90 120
Total reflections	30926 (3047) - 2.75)
Unique reflections	15464 (1524)
Multiplicity	2.0 (2.0)
Completeness (%)	99.23 (99.15)
Mean I/sigma(I)	23.39 (1.41)
Wilson B-factor	101.38
R-merge	0.01269 (0.5287)
R-meas	0.01795 (0.7477)
R-pim	0.01269 (0.5287)
CC1/2	1 (0.78)
CC*	1 (0.936)
Reflections used in refinement	15455 (1523)
Reflections used for R-free	777 (72)
R-work	0.2976 (0.5009)
R-free	0.3220 (0.5591)
CC(work)	0.909 (0.678)
CC(free)	0.963 (0.589)
Number of non-hydrogen atoms	3906
macromolecules	3903
ligands	3
Protein residues	534
RMS(bonds)	0.012
RMS(angles)	1.51
Ramachandran favored (%)	92.42
Ramachandran allowed (%)	7.39
Ramachandran outliers (%)	0.19
Rotamer outliers (%)	0
Clashscore	18.63
Average B-factor	137.73

The vast majority of BMC shell proteins assemble into sheets when crystallised. These include proteins closely related to *C. phytofermentans* PduT, such as PduT from *C. freundii* and *S. typhimirium*, as well as the PduT like protein from *H. ochraceum* [20, 24, 85]. Non- tessellating mutants have previously been described (chapter: 3), but a non- tessellating WT BMC shell protein is currently unique. Why PduT from *C. phytofermentans* does not tesselate is not immediately obvious. When comparing the sequence of both BMC domains from PduT *C. phytofermentans* to PduA and both PduT domains from *C. freundii* PduT (figure: 4.2A), the residues involved in tessellation are largely conserved. It is necessary to compare both BMC domains within a single PduT molecule as these domains may differ slightly in their overall sequence. Consequently, residues involved in oligomerisation may only be conserved in a single domain, yet this is still sufficient to promote tessellation. The interfacial lysine is fully conserved across all orthologues. The interfacial valine (which promotes hydrophobic packing) is partially conserved in the form of Leucine in domain 1 and isoleucine in domain 2 (both of these hydrophobic residues should function in a similar manner to valine in promoting tessellation). However, the primary difference arises when comparing the conservation of the arginine present at position 79 in PduA. This residue is only conserved in the second BMC domain of PduT (*C. freundii*) and neither of the BMC domains in PduT (*C. phytofermentans*).

To better understand the implications of the arginine in tessellation, the manner in which PduT (*C. freundii*) tessellates must be examined. Here, the interfacial lysine does not actually play a role. The side chain amine and carboxyl groups of the anti- parallel lysines are 6.9 Å apart, and so are unlikely to contribute to the interaction. Whilst the arginine in chain one is not conserved, the one in chain two is (arginine 172). This arginine is able to form salt bridges with the carboxyl groups of lysine 115 and serine 117, thus facilitating tessellation (figure: 4.2B). Due to the repeating nature of PduT (where two BMC domains are present in a single chain), conservation of only one of these arginines is sufficient to promote tessellation. In PduT (*C. phytofermentans*) the arginine is not present in either domain and the lysines are similarly distant, so as not to facilitate an interaction. It is therefore the absence of this arginine which accounts for the lack of tessellation seen in PduT

(*C. phytofermentans*).

It is unusual that the lysines are conserved in all structures examined, and indeed throughout most BMC shell proteins, but do not seem to contribute to tessellation in the case of either PduT orthologue. However, whilst they may not be suitably proximal to allow for tessellation with like shell proteins, they may be suitably located to allow for tessellation with other shell proteins. This is evidenced in the structure of a whole microcompartment [85]. Here, despite the PduT like proteins tessellating in the crystal structure in the absence of any other shell protein, they interacted exclusively with PduA like proteins when present in the whole microcompartment.

A

```
PduA_C.freundii                MQQEALGMVETKGLTAAIEAADAMVKSANVMLVGYEKI--GSGLVTVIVRGDVGAVKAAT 58
PduT_C.freundii_BMC_1          -MSQAIGILELTSIAKGMEAGDAMLKSANVNLLVSKTI--CPGKFLLMLGGDVGAVQQAI 57
PduT_C.freundii_BMC_2          --------VETWSVAACICAADRAVKASNVTLVRVHMAFGIGGKCYMVVAGDVSDVNNAV 52
PduT_C.phytofermentans_BMC_1 -MSKAIGMVEYKTVSSGIMAADLMVKTADIDIVEAQTV--CPGKYIILITGDLSAVNASV 57
PduT_C.phytofermentans_BMC_2 --IEALGVLETYSAASIIVAADVAAKTAAVELVEVRVARGMCGKSYLMLTGEIASVTASI 58
                                        :*    :  : *.*   *:: : ::  .      *   ::: *::. *  :

PduA_C.freundii                DAGAAAARNVGEVKAVHVIPRPHTDVEKILPKGIS--------- 93
PduT_C.freundii_BMC_1          ATGTSLAG--DMLVDSLVLPNIHASVLPAISGLNSVDKRQAVGI 99
PduT_C.freundii_BMC_2          TVASESAGEKGLLVYRSVIPRPHESMWRQMVEG----------- 85
PduT_C.phytofermentans_BMC_1 EAARIQFE--THLIDSFILGNPHDGILPAIYGASVVEE------ 93
PduT_C.phytofermentans_BMC_2 EAAKKAIGENGMYLDSSVLAHPDKKLRNKLM------------ 89
                              ..             :: . .  :   :
```

B

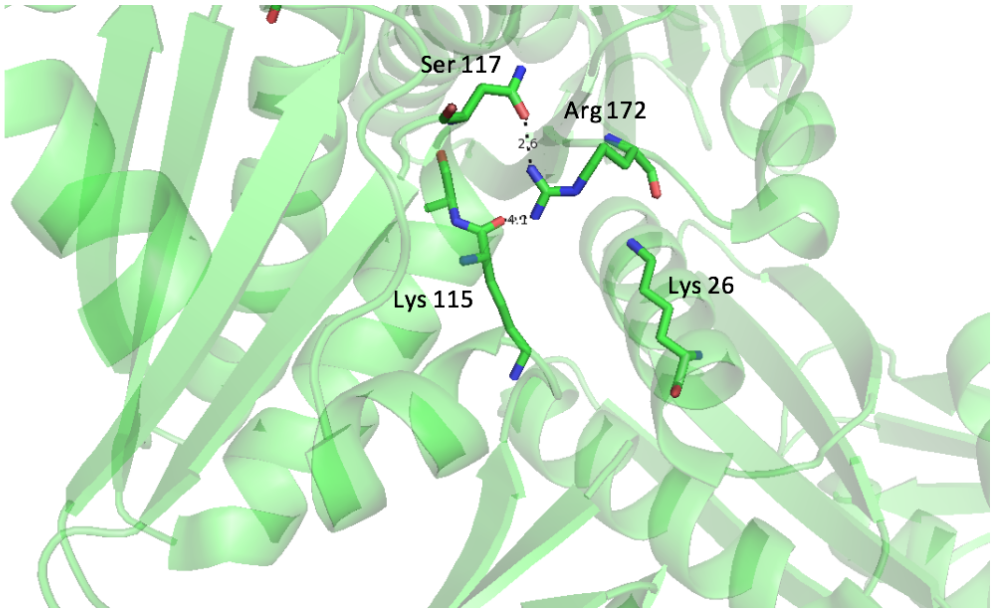

FIGURE 4.2: Alignment of PduA and each BMC domain from PduT (Both *C. freundii* and *C. phytofermentans*)(A), showing the absence of the arginine implicated in tessellation. The conservation of this arginine in only a single domain within PduT (*C. freundii*) is sufficient to promote tessellation. The position of this arginine in the second domain of PduB (*C. freundii*) (B) is shown to demonstrate how this residue facilitates tessellation. Arginine 172 in one pseudohexamer is only 2.6 Å and 4.1 Å apart from the carboxyl groups of serine 117 and lysine 115 in the adjacent psudohexamer. These interactions are sufficient to promote tessellation.

4.3 PduT features a tetrahedral metal binding site at the interface between two trimers.

During refinement, a region of unoccupied density was observed on the edge of a single PduT trimer. This density was proximal to two negatively charged residues,

glutamate 95 and histidine 172. The proximity to negatively charged residues inferred the presence of some cationic species. When viewing the symmetry mates it was clear that the same two residues from an adjacent trimer were also involved, and that there was a tetrahedral binding site at the interface of two trimers. There were no cations capable of interacting with histidine and glutamate in such a manner in the crystallisation solution (20 % PEG 8K, 100 mM TRIS pH 8.5, 200 mM Sodium Acetate). Therefore, this cation must have co- isolated with the protein during purification, or as it was expressed in the bacteria. The tetrahedral configuration with two glutamates and two histidines is remarkably similar to previously described Zn^{2+} binding sites. When Zinc was inserted into the position and refined, the positive density disappeared indicating it was potentially correct (figure: 4.3).

It is doubtful this Zn^{2+} binding site is biologically relevant and is more likely an artefact of the crystal as indicated by analysis of the interfaces with PISA [88]. In the context of the crystal the Zinc does appear to have a stabilising influence and may contribute to the stability and formation of the crystal.

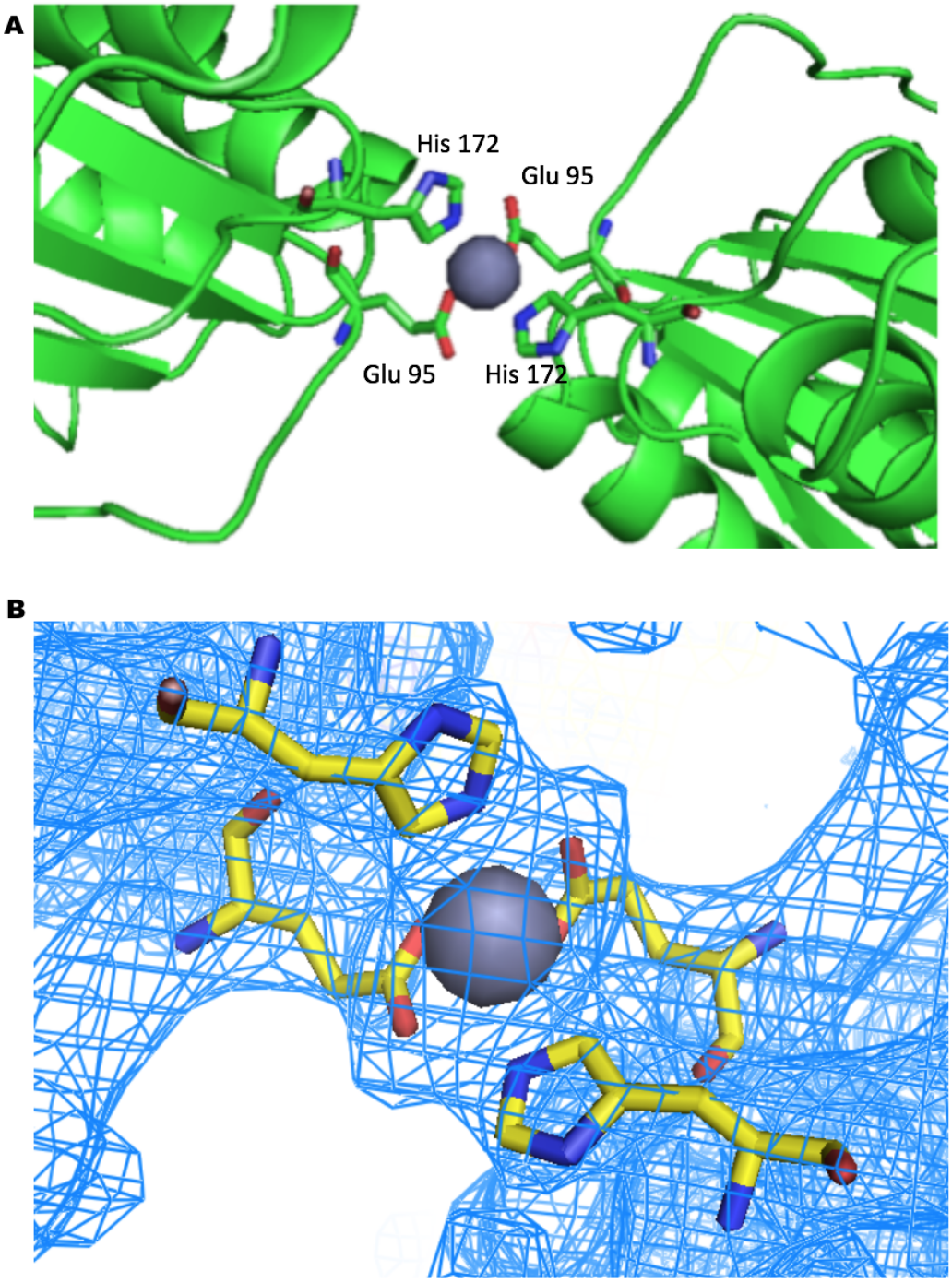

FIGURE 4.3: The proposed tetrahedral metal binding site at the interface of two PduT pseudohexamers (A), with glutamate 95 and histidine 172 being contributed from each PduT molecule. An (F(o) - F(c)) map is also shown, contoured to 3.0 sigma level (B), displaying the density corresponding to the Zinc atom.

4.4 PduT features an oxygen labile iron sulphur cluster which can be modelled in the pore.

The *C. phytofermentans* PduT was initially selected as a protein of interest by colleagues at the university of Kent due to its having two cysteines proximal to the pore of the trimer (at positions 38 and 133). PduT is known to feature an Fe- S cluster, as determined by SPR and the intense brown colour seen in purified proteins and crystals alike [24]. This cluster has previously been modelled in the pore region of *C. freundii* PduT [20], which only features a single cysteine per monomer at position 38. This results in one of the irons in the 4Fe- 4S cluster being free and potentially leading to the instability of the structure. Consequently, this explains why these clusters are usually not detectable in the diffraction data. It was speculated that by PduT (*C. phytofermentans*) having two cysteines per monomer (and therefore six in the whole trimer) that the 4Fe- 4S cluster may be fully co- ordinated and more stable. In spite of this, no density was observable in the pore of the crystal structure and the positions of cysteines at positions 133 were not suitably located to bind an 4Fe- 4S cluster. However, the cysteines at position 38 pointed inwards and superimposing an 4Fe- 4S cluster into the pore of the trimer (figure: 4.4), demonstrated that these cysteines are likely the residues which bind the 4Fe- 4S cluster, contributing to the brown colouration observable. However, this structure is not stable enough to be detected using X- ray crystallography.

In the context of the larger microcompartment, PduT likely exists in a complex with some internalised protein, which in turn contributes a cysteine to the 4Fe- 4S cluster and stabilises it. Until the identity of this complex is known, obtaining structural information via crystallography on iron- sulphur clusters within BMC shell proteins seems unlikely due to their instability.

FIGURE 4.4: A 4Fe- 4S cluster modelled in the pseudohexamer pore, the position of three pore lining cysteines at position 38 have suitable geometry for the insertion of a 4FE- 4S cluster.

4.5 Conclusions and further work.

The lack of tessellation seen in the *C. phytofermentans* PduT crystal structure is currently unique. The implication of this is that, PduT from *C. phypofermentans* must interact with some other shell protein in the complete BMC. The identity of this shell protein is unknown. However, it could be identified using pull down assays with other proteins from the *C. phypofermentans* Pdu BMC. Identification of which shell protein interacts with PduT would be beneficial for determining which shell proteins interact with which others, thereby illuminating how BMCs might assemble.

The identity of the metal binding site at the edge of the PduT trimers (section: 4.3) is not certain. Whilst it seems highly likely that it is zinc, this could be confirmed by an anomalous difference Fourier map. Here, where the incident X- ray has a wavelength close to the transition energy of the atom, the atom will be excited into a higher energy state [89]. Therefore, using an incident X- ray close to the transition energy of zinc would allow for confirmation that this is the metal bound.

The lack of a resolvable iron- sulphur cluster in *C. phytofermentans* PduT is consistent with previous PduT structures from *C. freundii* and *S. typhimirium* [20, 24].

Unfortunately, the cysteine present at position 133 in *C. phytofermentans* PduT was not suitably proximal to the pore, to bind a 4Fe- 4S cluster. As has been speculated with other PduT orthologues, it is likely that some internalised enzymes contributes the final cysteine to bind the 4Fe- 4S cluster. Identification of this enzyme and co-crystallisation with PduT may stabilise the 4Fe- 4S cluster so that the cluster is resolvable.

Chapter 5

An investigation into the internalised enzyme, PduL.

5.1 Introduction.

PduL is a phosphotransacetylase (PTAC) found in the Pdu BMC. The function of PduL is to transfer a phosphate ion to propionyl- CoA, forming propionyl phosphate and CoA. CoA can then be used by PduP to form more propionyl- CoA and so the actions of the two enzymes are tightly linked. The majority of enzymes in the Pdu BMC are homologous to enzymes found in the cytoplasm but feature an N- terminal extension which facilitates their encapsulation. PduL is unique in this regard and is exclusive to the Pdu BMC, having no cytoplasmic equivalent. As with other encapsulated enzymes PduL features an N- terminal extension (L20), promoting its encapsulation.

The purpose of this chapter is to investigate the structural characteristics of PduL, particularly in regard to the N- terminal signalling sequence.

5.2 Purification and preparation of PduL.

PduL from *C. freundii* was produced in the method outlined in chapter 2 (section: 2.3.5). Analysis with SDS- PAGE (figure: 5.1) revealed a single band at ~26 kDa following Nickel IMAC, in line with the theoretical weight. PduL eluted from

an S200 column at 14.8 ml, a result consistent with PduL being a dimer. This is consistent with PduL from *S. enterica* [90], which shares 87 % sequence identity with *C.freundii*. PduL features an N- terminal signalling sequence. Consequently, the N- terminal his- tag was cleaved to better expose the signalling sequence. This was achieved through overnight incubation with thrombin at 4 °C, with cleavage being confirmed by SDS- PAGE. Samples exposed to thrombin displayed a slight decrease in the molecular weight (~2 kDa, or approximately the length of a his- tag with linker sequence). Following all purification steps, PduL was produced to a high homogeneity.

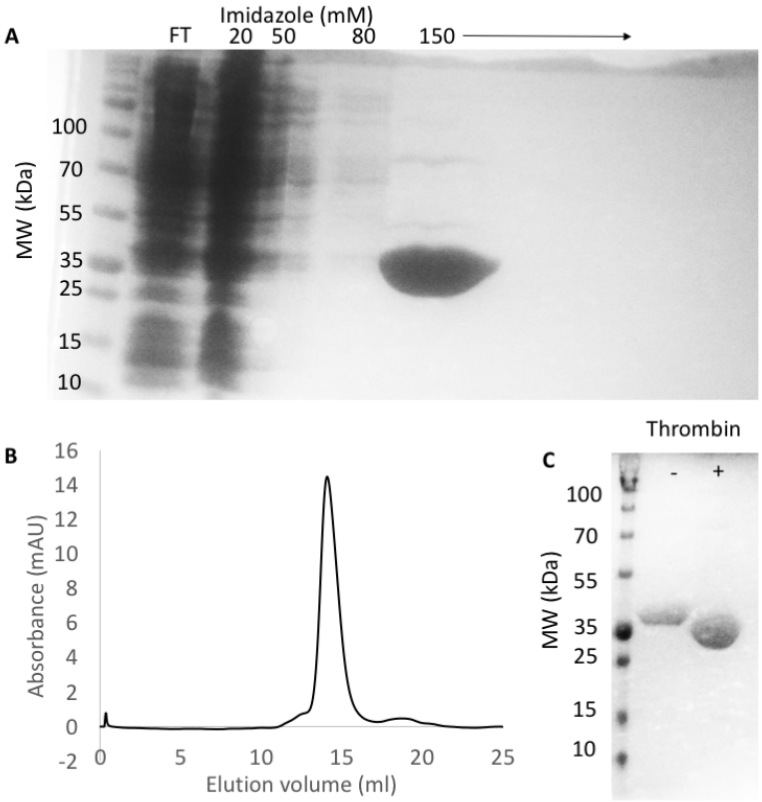

FIGURE 5.1: Purification of PduL. (A) SDS- PAGE of IMAC fractions. PduL elutes following the addition of 100 mM imidazole at a mass of 26 kDa, consistent with the theoretical mass. (B) SEC of the purified sample using an S200 column, yielded a single peak at 14.8 ml, consistent with the mass of a dimer (~46 kDa). (C) SDS- PAGE of the purified sample before and after treatment with thrombin. The small decrease in mass following incubation with thrombin is consistent with the mass of a his- tag, indicating cleavage of the tag was successful.

5.3 Structural characteristics of the PduL signalling peptide.

Circular dichroism is a powerful tool for analysing the secondary structure of proteins and changes that arise in these structures following binding events. The premise of CD involves measuring the differential amounts of circularly polarised light absorbed by chiral molecules. Peptide bonds within proteins are chiral and exhibit changes in absorption based on local conformations within the molecule. α- helices exhibit absorption maxima at 208 and 222 nm, β- sheets at 216 nm and random coils at 190 nm. This information can be used to assign structural features to the whole protein.

The CD spectra of the L20 peptide was compared to the signalling peptides of PduP and PduD (P18 and D18). P18 is known to be α- helical from NMR data [29]. However, the structural characteristics of L20 and D18 are unknown. The CD spectra produced (figure: 5.2) shows that as expected, the P18 peptide is α- helical in conformation and D18 is also α- helical. L20 differs substantially and the dominant peak at 190 nm is indicative of a predominantly unstructured conformation. Why the structure of L20 should be so different to D18 and P18 is not immediately clear. All three signalling sequences share a high degree of sequence identity and feature the same hydrophobic motif (figure: 5.3). It is possible that some other element of PduL interacts with the signalling peptide in the whole structure, facilitating a different conformation compared to the peptide in isolation.

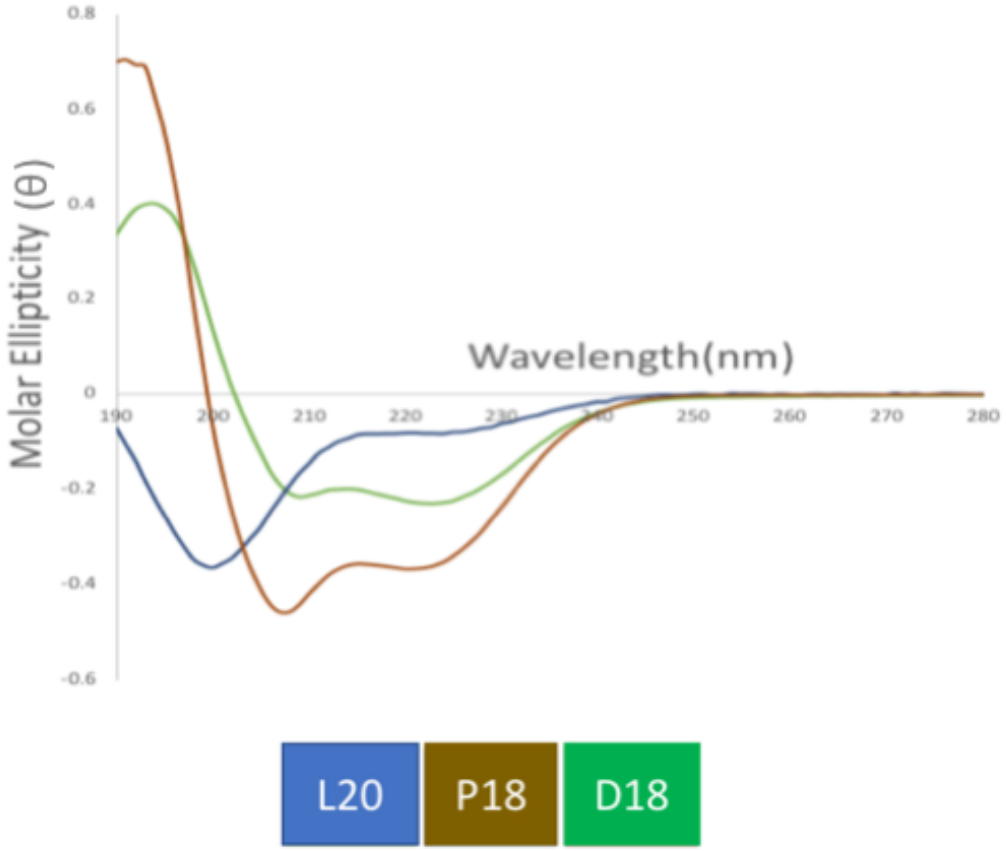

FIGURE 5.2: The CD spectra of D18 (green), P18 (brown) and L20 (blue). P18 and D18 are seen to be clearly alpha helical with peaks at 208 and 222 nm whereas L20 is largely disordered, displaying a prominent peak at 190 nm.

```
PduP --MNTSELETLIRNILSEQL-- 18
PduL --MDKQQLETTVAKVLDEMRER 20
PduD MEINEKLLRQIIEDVLSEMQTS 22
       ••   *   •   •* *
```

FIGURE 5.3: Alignment of known Pdu signalling peptides. Fully conserved residues are highlighted green and partially conserved residues are highlighted blue. All three peptides feature a conserved hydrophobic motif.

5.4 Crystal structure of PduL.

To better understand PduL and its N- terminal signalling peptide, X- ray crystallography was utilised. PduL from *R. palustris* had previously been crystallised (PDB: 5CUO) [90]. However, this PduL orthologue was relatively unstable and required the removal of 40 N- terminal residues to facilitate crystallisation. In contrast PduL from *C. freundii* is stable and shows little aggregation. Therefore, crystallisation of the complete protein was attempted. The JCSG, Structure screens 1 and 2 and morpheus products from molecular dimensions were used to screen crystal conditions. Crystals were observable after approximately 5 days (figure: 5.4) under the following conditions: 8 mg/ml PduL, 0.2 M lithium sulphate, 0.1 M Sodium Acetate, 50 % PEG400.

5.4.1 Data collection and reduction.

Data were collected at Diamond Light Source on beam line I04 and X- ray wavelength used for data collection was 0.97628 Å. Data collection was successful with data being collected to a maximum resolution of 1.84 Å. Data were then reduced and integrated using Xia2/3dii suggesting a space group of P3121, completeness of 99.93% and multiplicity of 2.0. These statistics were all suitable to continue with attempts at structure elucidation.

The PHENIX software suite was used for all following steps. Molecular replacement was attempted using PduL from *R. palustris* (PDB: 5CUO) [90] as a search model, which has a relatively high sequence identity with *C. freundii* PduL (38.2%). Molecular replacement was successful yielding a TFZ score of 22.6 (it is generally accepted that a TFZ score of 8.0 or above is required for further refinement and validation to be permissable). Further refinement was performed using the refine tool within PHENIX and elements of the structure rebuilt using Coot. Structure validation using Molprobity was completed once an Rfree value of 0.25 had been attained.This produced a score of 1.40 (ideally the score should be lower than the maximum resolution of the data, which in this case was 1.84). There were two copies of PduL per

asymmetric unit in the final structure.The final refinement statistics are summarised in table:

TABLE 5.1: Refinement statistics of the completed PduL structure and crystallographic data.

Wavelength	0.9795
Resolution range	**59.34 - 1.84 (1.906 - 1.84)**
Space group	P 31 2 1
Unit cell	76.6258 76.6258 132.528 90 90 120
Total reflections	79524 (7814)
Unique reflections	39762 (3907)
Multiplicity	2.0 (2.0)
Completeness (%)	99.93 (99.80)
Mean I/sigma(I)	12.59 (1.25)
Wilson B-factor	31.61
R-merge	0.02984 (0.5149)
R-meas	0.0422 (0.7282)
R-pim	0.02984 (0.5149)
CC1/2	0.999 (0.667)
CC*	1 (0.894)
Reflections used in refinement	39739 (3901)
Reflections used for R-free	1864 (185)
R-work	0.1936 (0.2985)
R-free	0.2387 (0.3445)
CC(work)	0.958 (0.781)
CC(free)	0.949 (0.718)
Number of non-hydrogen atoms	3293
macromolecules	2951
ligands	52
solvent	290
Protein residues	386
RMS(bonds)	0.013
RMS(angles)	1.49
Ramachandran favored (%)	96.86
Ramachandran allowed (%)	3.14
Ramachandran outliers (%)	0
Rotamer outliers (%)	0
Clashscore	10.12

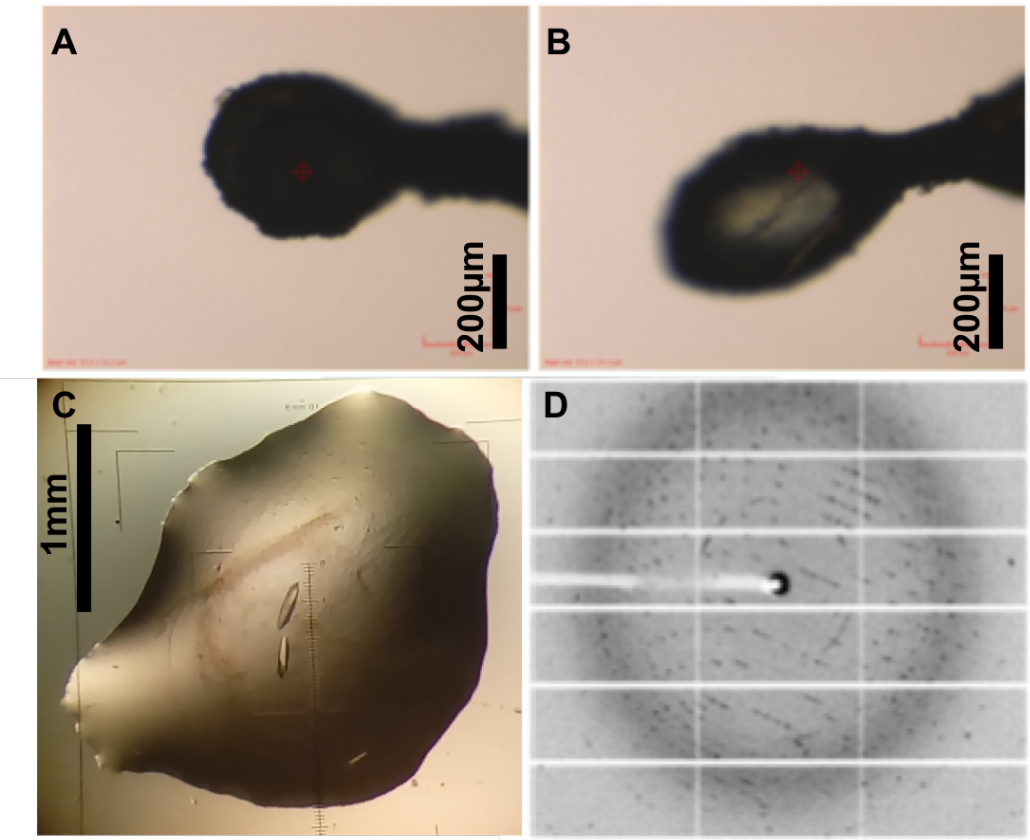

FIGURE 5.4: PduL crystal mounted on a 200 (μm litholoop, imaged at diamond light source (A,B). Image of PduL crystals as they appeared in the drop (C). Typical diffraction data collected from PduL crystals (D).

5.4.2 The N- terminal signalling peptide of PduL is unstructured in the crystal.

Despite full length PduL being crystallised, the N- terminus is unstructured in the crystal. Residues from 14 onwards are clearly visible (figure: 5.5) in chain A and residues from only 20 onwards are visible in chain B. This is contrary to popular belief that the signalling peptides from PduP, PduD and PduL form amphipathic alpha helices, but is in accordance with CD experiments outlined previously (figure: 5.2).

The asymmetry between chains A and B is accounted for by the crystal contacts formed between Arg 18 in Chain A, and Asp 164 in chain B of an adjacent symmetry mate (figure: 5.6). This short ordered segment of the signalling peptide from 14- 20

is therefore likely an artefact of the crystal, rather than being of any biological relevance. The arrangement of molecules within the lattice is such that the N- terminus of chain B is not proximal to other symmetry related molecules. Therefore, without any interactions stabilising the N- terminus, no structure is determinable. Figure: 5.7 demonstrates the manner in which the N- terminus of chain A extrudes into the adjacent molecule. This anchors it in position and allows for this short ordered sequence, which would likely not exist in solution.

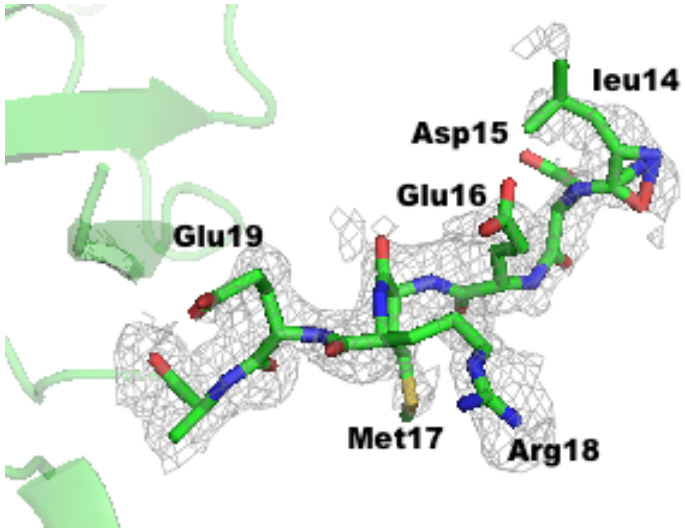

FIGURE 5.5: Map of the PduL N- terminus (2F(o) - F(c)) contoured to 2.0 sigma level. A lack of density clearly showing the first 14 residues are disordered. However, residues 14- 20 show clear density, suggesting these residues are ordered.

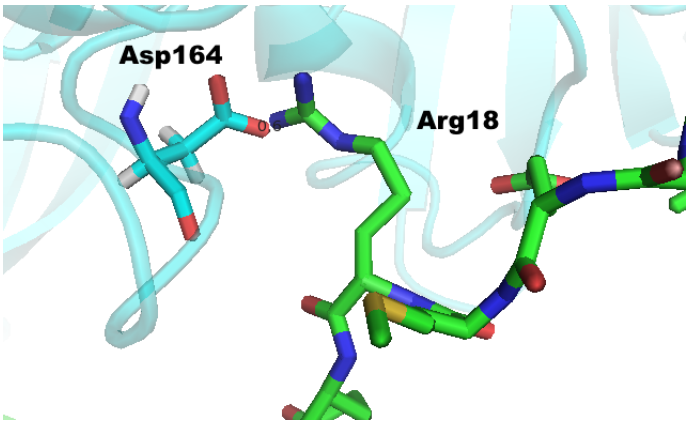

FIGURE 5.6: Inter- molecular interactions within the PduL crystal structure. An interaction between Arg 18 in Chain A and Asp 164 in chain B of an adjacent symmetry mate stabilises the N- terminus, allowing for the fitting of residues 14- 20.

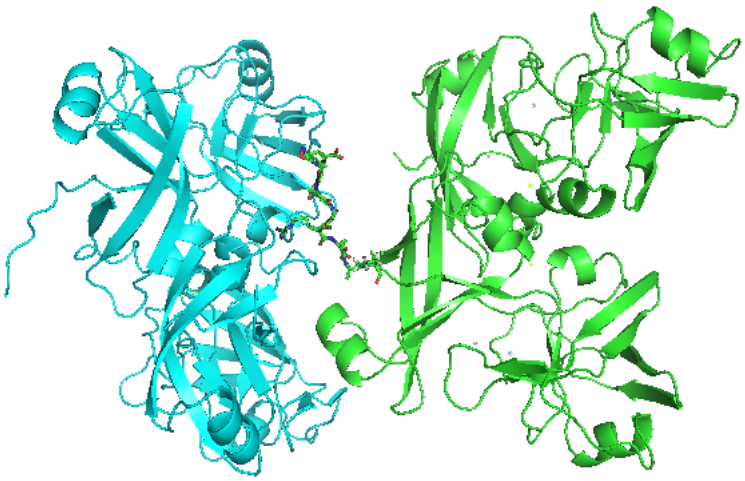

FIGURE 5.7: Representation of the packing of two PduL symmetry mates, the N- termini of one PduL (green) is stabilised by the placement of its N- terminus with respect to another PduL present in the lattice (blue).

5.4.3 Only a single active site is occupied by Coenzyme A in the PduL dimer.

In addition to the asymmetry observed between the N- termini of chain A and B in the crystal structure, only the density for a single CoA molecule was observable in the active site of chain A. No density was apparent in the active site of chain B. This is despite both molecules being identical sequentially and having an RMSD of only 0.423 Å when superimposed on one another. This slight difference in RMSD is primarily a product of shifts in positions of residues within the CoA binding site, which undergo a conformational change following the binding of CoA.

The reason for no CoA being bound to chain B is unclear. Each CoA occupies a unique site and so there is ample scope for two CoA molecules binding. It may be that PduL can only bind one CoA at a time (although why this should be is not clear from the crystal structure). However, this is unlikely the case for PduL as its catalytic mechanism is reasonably well documented [50]. Sufficient CoA was added to the crystallisation mix (4:1 molar ratio) to provide a ratio of CoA to PduL for total saturation. However, this has not occurred. There have been several documented cases of active site asymmetry in crystal structures with no apparent cause [91, 92].

There are two tetrahedral metal binding sites within the PduL dimer (one per monomer). These sites are co- ordinated entirely by histidines (figure: 5.9). The crystal structure of PduL from *R. palustris* featured a Zn^{2+} binding site at the same location [90] and so it is highly probable that this is also a Zn^{2+} binding site. As the function of PduL is to hydrolyse Propionyl- CoA, Zn^{2+} likely participates as a lewis acid in this reaction.

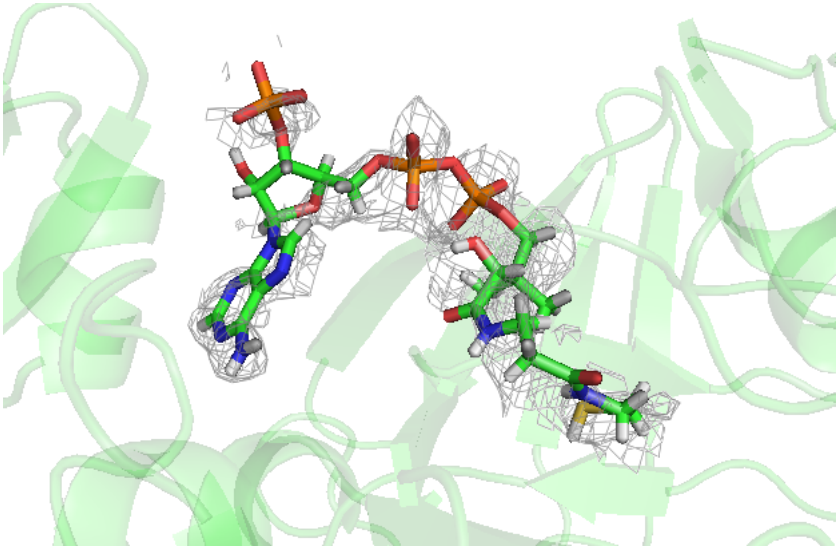

FIGURE 5.8: A single CoA molecule is bound to PduL. An F_{Obs} - F_{Calc} map shown here clearly shows the presence of CoA in one binding site. There is particularly strong signal at the phosphates within the CoA molecule, a product of this region binding to positively charged residues and so is essentially anchored in place. The other binding site features no such density and is therefore presumably unoccupied.

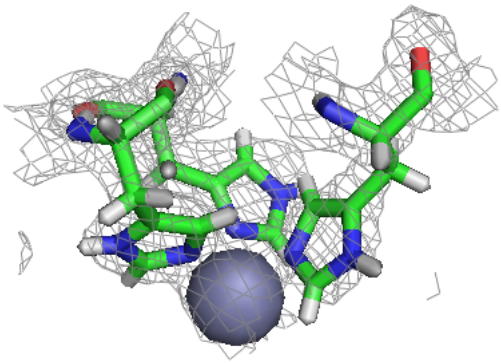

FIGURE 5.9: F_{Obs} - F_{Calc} map showing the clear presence of a Zn^{2+} binding site, co- ordinated by histidines within the PduL active site.

5.4.4 Differences between PduL from *C. freundii* and *R. palustris*.

In order to highlight any structural differences between PduL from *C.freundii* vs *R.palustris*, a structural superimposition was performed using PyMOL. After alignment, the RMSD was only 0.825 Å; not particularly significant for two proteins which share almost 40 % sequence identity (figure: 5.10). Although almost all secondary structural elements were highly homologous, a loop at 196- 202 was the only region which was visually divergent (figure: 5.11). No residues in either structures from this region are involved in binding CoA. However, aspartate 198 in the *C.freundii* structure forms a salt bridge with arginine 96 in the other monomer within the PduL dimer. This salt bridge is not present in the *R.palustris* structure as neither aspartate 198 nor arginine 96 are conserved. Therefore, this interaction drastically alters the position of the loop between each structure.

The interaction between aspartate 198 and arginine 96 in the *C. freundii* structure appears to be one of relatively few interfacial interactions. These interactions act as contacts between the two monomers within the dimer and are therefore, important for maintaining the oligomeric state. Despite this, neither aspartate 198 or arginine 96 are conserved in the *R. palustris* structure. It is unusual for residues involved in oligomerisation to not be conserved. This is especially relevant as it had been speculated that phenyalanine 116, which is present at the same position as arginine 96 in the *C. freundii* structure was critical for the stability of the *R. palustris* PduL dimer [90]. It was also speculated that PduL orthologues which lacked this residue must therefore have some other oligomeric state. Phenylalanine 116 is also critical for CoA binding in *R. palustris* PduL. However, the substitution of phenylalanine for arginine in *C. freundii* may not have any large impact on the binding of CoA. Arginine is capable of forming a hydrogen bond with the Nx atom of the adenosine ring within CoA at pH values higher than 4. Therefore, hydrogen bonding could bind CoA to PduL in *C. freundii*, rather than the $\pi - \pi$ stacking present in the case of Phenylalanine interacting with CoA in *R. palustris*.

In light of the lack of conservation of residues critical for oligomerisation, the residues involved in dimerisation of *C. freundii* PduL were characterised and are shown in

table: 5.2. What is remarkable is that only three of the thirteen residues involved in the oligomerisation are fully conserved across both PduL orthologues (figure: 5.10). This raises questions regarding the stability of different PduL orthologues and how the binding of CoA effects their stability. Table: 5.3 contains the outputs from the PISA webserver, which calculates thermodynamic stability of PDB files. PduL from both *C. freundii* and *R. palustris* were provided as inputs, with both an empty active site and CoA bound. The output parameter ΔG_{diss} is the Gibbs free energy of dissociation. A negative ΔG_{diss} value indicates an unfavourable process, where some outside factor must be applied for the oligomer to remain stable. PduL from *R. palustris* has a slightly negative ΔG_{diss} where CoA is bound, at -1.8 J. When CoA dissociates this value decreases to -6.0 J. Therefore, CoA may function as a kind of anchor between the two chains, tethering them together by interacting with residues from both chains in the dimer. Where CoA is not bound, this may be sufficient to trigger dissociation. In contrast, the ΔG_{diss} for *C.freundii* is -6.4 J where CoA is bound, and -6.6 J where CoA is unbound. This reflects no substantial change in thermodynamic stability, although it is somewhat surprising to see even a marginal decrease in stability following CoA binding. PISA predicts both these structures to be unstable and that the dimer present in the crystal structure is an artefact of crystallisation. However, data from SEC (figure: 5.1) indicates PduL is a dimer. *C. freundii* PduL elutes from an S200 column at 16.8 ml, which equates to a molecular weight of 44.151 kDa, or slightly lower than the anticipated weight for a PduL dimer (45.36 kDa).

The effects on the stability of the *C. freundii* PduL dimer induced by the absence of phenylalanine 116 were explored. Arginine 96 in the *C. freundii* PduL structure was mutated insilico to phenylalanine. This mutation did have a substantial effect on the stability of the dimer, increasing the ΔG_{diss} to -6.4 from -4.1 J. Although this is not as stable as the *R. palustris* dimer, it is clear that phenylalanine 116 plays a crucial role in stabilising the PduL dimer in *R. palustris*. Phenylalanine 116 was also mutated to arginine in the *R. palustris* structure. Here, the inverse effect was apparent where stability decreased from -1.8 to -2.9 J, highlighting the significance of this residue in PduL dimer stability.

```
C.freundii   --MDKQQLETTVAKVLD----------------EMRERPIPLGISNRHIHLCAEDYGRLF 42
R.palustris MDMQQETIERIIRQVLGQVGPAGGSIATLSGDAGVDPFQVAVGVSNRHIHLSRTDMDTLF 60

C.freundii   P-NHPISEKKELLQPGQYAAEQTVTLVGPKGQLKNVRLLGPLRNTSQVEISRTDARTLGI 101
R.palustris GPGAELQRKKAMKQPGQFAAEETVTLKGPKGSLSKVRVLGPLRRETQVEVSVADGFALGI 120
                                                                  116
C.freundii   AAPLRMSGNIQGTPGVRLVSPFAELDLASGVIVAQRHIHMSPLDALILRVSHGDKVSVAI 161
R.palustris TPPLRQSGQLDDTPGLTIIGPQGSVTKDHGVIVAQRHIHMHPSTAAKLGLRNGDEVDVEA 180

C.freundii   NGDERRLIFDNVAVRVSPDMRLEMHIDTDEANAAGADNPQAFATLVTSR 210
R.palustris GGERG-GVMHRVLIRVAEASADEMHIDVEEANALCLKNDDVVRICKK-- 226
```

FIGURE 5.10: Alignment of PduL from *C. freundii* and *R. palustris*. Key residues which have been identified from the crystal structure as being involved in inter- subunit interactions are highlighted. Those residues which are fully conserved between both species are shown in green. Residues that are partially conserved are shown in yellow. Residues that are not conserved are highlighted in red. Phenylalanine 116 is not conserved and is shown at position 116.

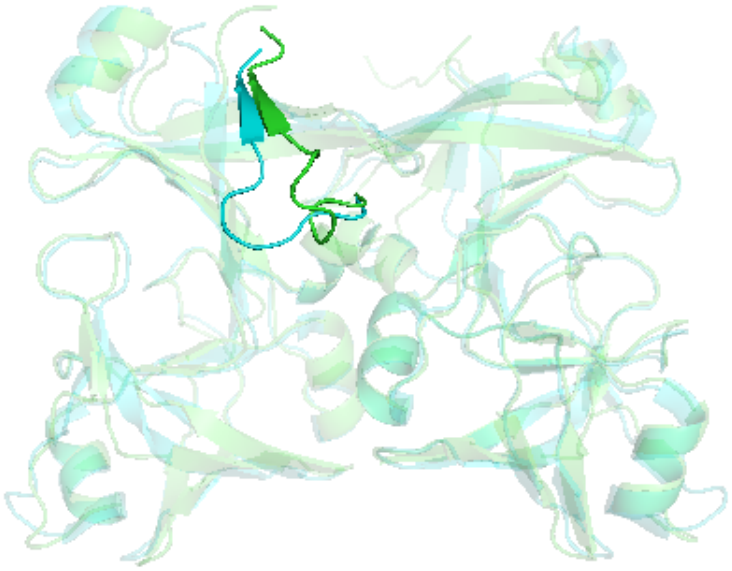

FIGURE 5.11: Structural alignment of PduL from *C. freundii* and *R. palustris*. *C. freundii* is shown in green and *R. palustris* in blue. The majority of the structures are highly homologous. However, a loop at position 196- 202 in *C.freundii* is highly divergent and is highlighted compared to the rest of the structure.

TABLE 5.2: The inter-chain electrostatic interactions identified as being involved in dimerisation of *C. freundii* PduL from the crystal structure. In *C. freundii* there are a large number of inter- chain interactions, not involving CoA.

Atom Chain A	Atom Chain B	Distance	Type of interaction
Thr 94 (OG1)	Asn 193 (HD22)	2.2	Hydrogen Bonding
Arg 97 (NH1)	Asp 198 (OD1)	2.7	Salt Bridge
Ala 135 (O)	Arg 166 (HH21)	2.3	Hydrogen Bonding
Arg 137 (HH22)	Arg 167 (O)	2.0	Hydrogen Bonding
Arg 137 (HH12)	Ala 194 (O)	2.2	Hydrogen Bonding
Arg 137 (HH22)	(Arg 166 (O)	2.5	Hydrogen Bonding
Glu 165 (O)	Asn 172 (HH22)	2.0	Hydrogen Bonding
Arg 166 (NH2)	Asp 188 (OD2)	2.8	Salt Bridge
Arg 166 (HH21)	Ala 135 (O)	2.2	Hydrogen Bonding
Arg 167 (NE)	Asp (171 OD1)	2.8	Salt Bridge
Arg 167 (O)	Arg 137 (HH22)	2.2	Hydrogen Bonding
Ile 169 (O)	Ile 169 (H)	2.0	Hydrogen Bonding
Ile 169 (H)	Ile 169 (O)	2.0	Hydrogen Bonding
Asn 172 (HD21)	0.Glu 164 (O)	2.0	Hydrogen Bonding
Asp 188 (OD2)	Arg 166 (NH2)	2.9	Salt Bridge
Asn 193 (OD1)	Arg 97 (HH12)	2.6	Hydrogen Bonding
Asn 193 (HD22)	Thr 94 (OG1)	2.3	Hydrogen Bonding
Asp 198 (OD1)	Arg 98 (NH2)	2.8	Salt Bridge

TABLE 5.3: Thermodynamic properties of PduL assemblies, It had been noted that Phe 116 was important for dimerisation in *R. palustris* however was not present in *C. freundii*. Therefore, the stability of the two proteins was assessed. Lower values of ΔG_{diss} represent a decrease in stability. In the *R. palustris* structure CoA binding results in an increase in the stability of the complex, however, the inverse is true for *C. freundii*.

PduL type	ΔG_{diss}(Joules)
PduL *C.freundii* (CoA bound)	-6.4
PduL *C.freundii*i (CoA unbound)	-6.6
PduL *R.palustris* (CoA bound)	-1.8
PduL *R.palustris* (CoA unbound)	-6.0
PduL *C.freundii*R96F (CoA bound)	-4.1
PduL *R.palustris* F116R (CoA bound)	-2.9

5.5 The oligomeric state of PduL fluctuates as a result of CoA binding.

Native PAGE analysis of PduL was performed with the addition of its substrates and cofactors: CoA, PO_4^{-2} and Zn^{2+}. This demonstrated that the addition of CoA alone was sufficient to significantly alter the migration of PduL (figure: 5.12). The reduction in migration following the addition of CoA is potentially a result of the dissociation of the PduL dimer into PduL monomers. Alternatively, the presence of a -4 charge on CoA may be sufficient to alter the overall charge of the PduL dimer to such a degree, that it produces a significant increase in migration. In order to determine whether the change in migration is a product of dimer dissociation, or an increase in the net negative charge of the complex, SEC was performed in the presence and absence of CoA. As SEC is a technique which separates by size of the molecule, any change in charge should not effect retention time. The peaks produced following addition of CoA were at 18.5 and 14.8 ml vs a single peak at 14.8 ml where no CoA was present. This is consistent with the masses of a dimer and monomer. The values calculated for protein stability in the presence and absence of CoA (table: 5.3), combined with the Native PAGE and SEC data indicate that the binding of

CoA to *C. freundii* PduL is sufficient to trigger the dissociation of the dimer.

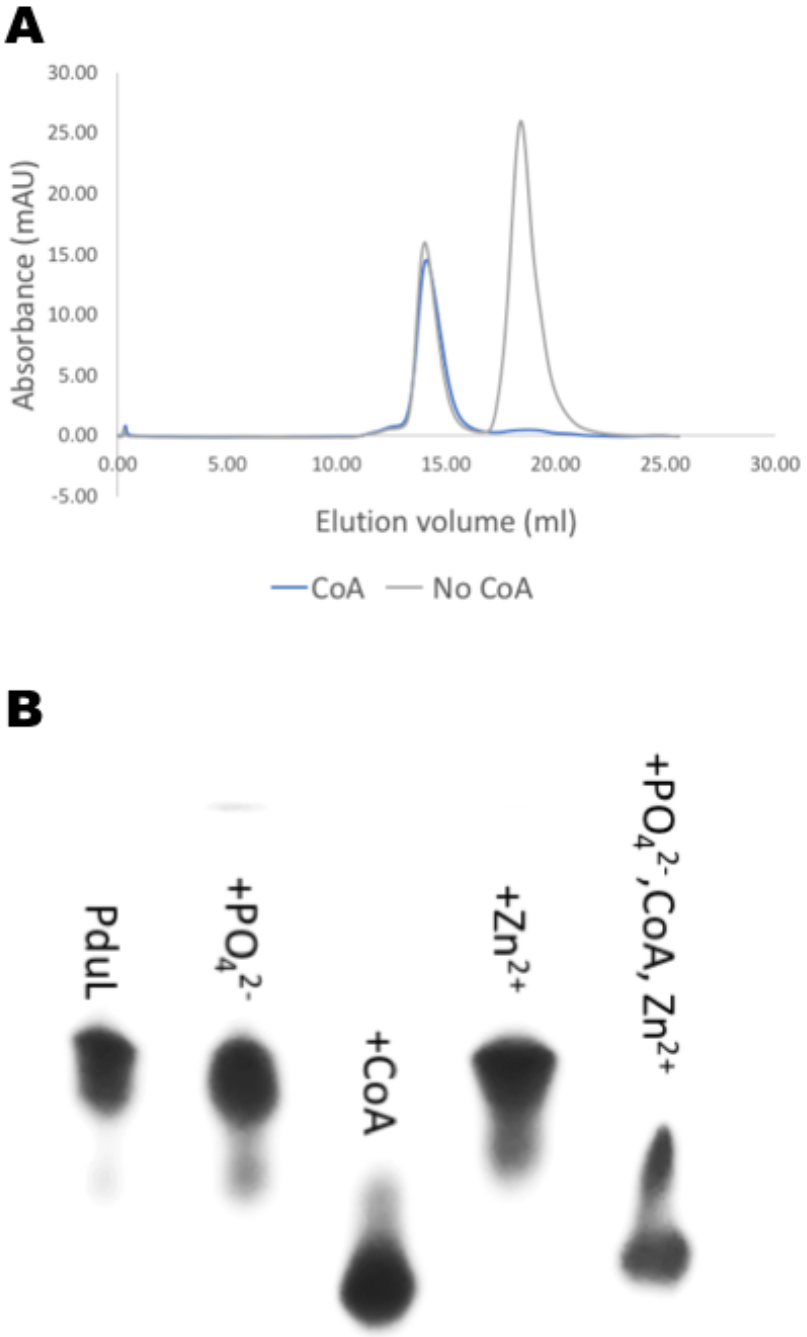

FIGURE 5.12: SEC profile of PduL in the presence and absence of CoA (A), the two peaks produced following addition of CoA implies the dissociation of the PduL dimer following CoA binding. Native PAGE (B) also shows an increase in PduL migration following addition of PduL, reflecting a decrease in mass, likely as a product of dissociation of the PduL dimer into monomers.

5.6 Conclusions and further work.

It was hoped that the crystal structure of complete PduL could be used for docking studies with shell proteins, in order to investigate the mechanisms of encapsulation. Despite this, evidence that the N- terminus of PduL is unstructured in both solution and in the crystal is in accordance with current thought on signalling peptides [93]. Due to the function of signalling peptides being to bind to other proteins, they have a greater tendency to aggregate than other peptides. It is thought that by maintaining a disordered state in the peptide itself and the regions flanking it, this aggregation is reduced. However, signalling peptides often undergo changes in their secondary structure following binding to their target. Therefore, this presents an opportunity to investigate structural changes undergone following PduL binding to shell proteins. This concept is explored in depth in chapter: 6.

The change in oligomeric state following binding of CoA to PduL is unusual. It is unclear what benefit the dissociation of the dimer might have. However, PduL is unusual in that the active site is contained entirely within a single monomer and there is no obvious advantage to it existing as a dimer vs. monomer. One explanation is that the dissociation into monomers does have some influence on activity and this is a feedback mechanism of sorts to control flux of substrates within the BMC. To investigate this, more kinetics data is required on PduL, specifically the effect of CoA concentrations on activity. A titration of CoA into PduL, combined with Native PAGE could also be used to ascertain the CoA concentration required to facilitate the dissociation.

Chapter 6

Interactions between shell proteins and signalling peptides.

6.1 Introduction.

The precise manner in which encapsulated enzymes are incorporated into the microcompartment lumen has remained unclear for some time. Several studies using mutagenic and computational techniques have produced models (outlined in section: 1.6). These models implicate the role of the PduA C- terminus in binding to the N- terminal signalling peptides present on a number of microcompartment associated enzymes. These signalling peptides have been identified as being present on the N- terminus of the enzymes PduP, D and L (P18, D18 and L20).

One key issue with the models proposed [51, 52] is the recent mounting evidence that the convex (and thus C- terminal side) of hexameric shell proteins are present on the exterior of the microcompartment [30, 31] (figures: 1.14 and 1.12). If previous models are correct, and the enzymes are associated with the C- terminus of PduA, they must therefore be displayed on the microcompartment exterior. This would negate the functions the microcompartment shell has been proposed to possess. These functions include, protecting the cellular interior from harmful aldehyde intermediates, and sequestering substrates at a higher concentration within the microcompartment lumen.

There are several modes of binding which have been proposed. These include, binding of the targeting peptide to the C- terminus of PduA [51, 52], and binding at either the cleft present between two PduA monomers, or two tessellating PduA hexamers [94]. The latter had been previously discounted as PduA has been shown to co-elute with PduP in pull down assays and it was presumed that PduA existed as single hexamers in solution at low concentrations. However, the work in the chapter: 3 has clearly demonstrated that PduA is capable of oligomerising at low concentrations.

The purpose of this chapter is therefore, to probe the interactions between shell proteins and the encapsulated enzyme, PduL.

6.2 Probing protein- protein interactions.

Previous work assessing the action of signalling peptides has largely focused on a reductive approach. Mutants of BMC enzymes, lacking the signalling sequences were produced and observed to no longer co- isolate with purified BMC's, thereby inferring the significance of the removed sequence. Whilst this approach is able to ascertain which enzymes are assossciated with the BMC, it provides no information regarding where they bind, and to what shell proteins. There have been some efforts to demonstrate specific interactions between shell proteins and enzymes, namely between PduK and PduP via tryptophan quenching fluorescence [29] and PduA and PduP via pull down assays [48]. The intention of this section is to better catalogue the interactions between extension peptides and shell proteins, thus highlighting features shared by shell proteins which interact with the same peptides and the differences in those that do not.

6.2.1 Peptides co- elute with PduA following SEC.

The interaction between three enzymes which are known to feature signalling sequences (PduD, PduP and PduL) and the shell protein PduA was investigated. Here, short peptides consisting of the N- terminal signalling sequence of PduD, PduP and

PduL (D18, P18 and L20) were produced by Generon ® and incubated with PduA for 30 minutes in a 10:1 molar ratio. SEC of this sample yielded an elution profile identical to that of PduA alone with a peak at ~14ml (figure: 6.1). Concentrating this fraction to ~1mg/ml and analysing masses present using MALDI- TOF yielded striking spectra clearly showing only two major species, PduA at ~14 kDa and each peptide at around ~2 kDa (figures: 6.2, 6.3, 6.4).

As the peptides are substantially smaller than PduA, it can be assumed that they elute at a much later point. However, the peptides contain no aromatic residues and so their retention time can not be quantified using A_{280} absorbance. Any peptide which co- elutes with PduA at 14 ml should be associated in some capacity. Therefore, it is apparent that each of the three extension peptides binds to PduA. P18 has a greater intensity relative to L20 which in turn is greater than D18. It is important to note that no inference can be made regarding affinity from the relative size of the peaks. Whilst more P18 may be present than either L20 or D18, the size and charge of the peptides also influences their flight and detection and so it is impossible to determine the amount of each peptide present without a standard curve.

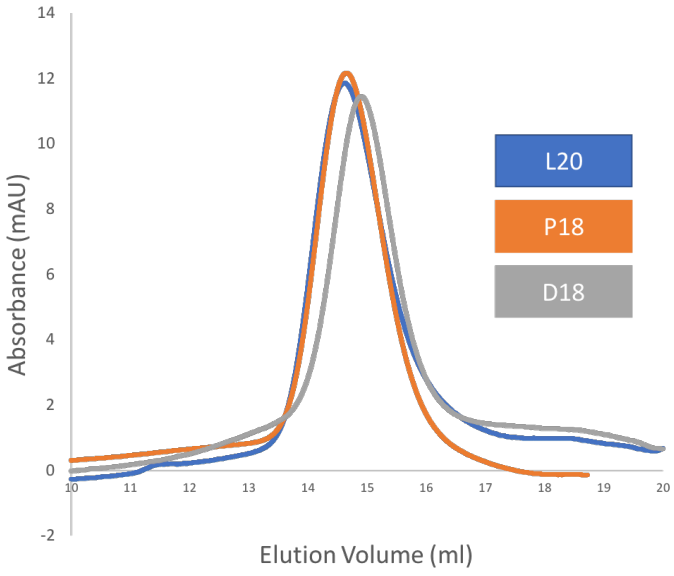

FIGURE 6.1: Elution profile of PduA with the addition of L20, P18 and D18 peptides. PduA eluted at the same point (~14.1 ml) regardless of which peptide was added. The elution volume of the peptides is not visible as none feature aromatic residues which absorb UV light at A_{280}.

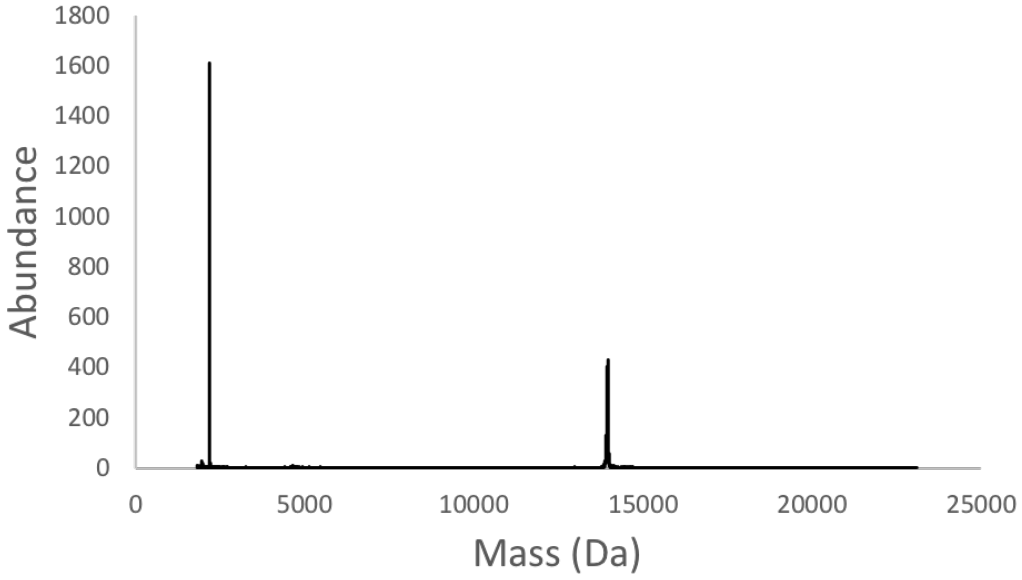

FIGURE 6.2: MALDI- TOF spectra of a single peak taken from the SEC of PduA and P18. Peaks at ~1800 and ~14,000 correspond to P18 and PduA respectively, indicating that both P18 and PduA are present in the sample and the two polypeptides have eluted at the same point, indicative of an interaction between the two.

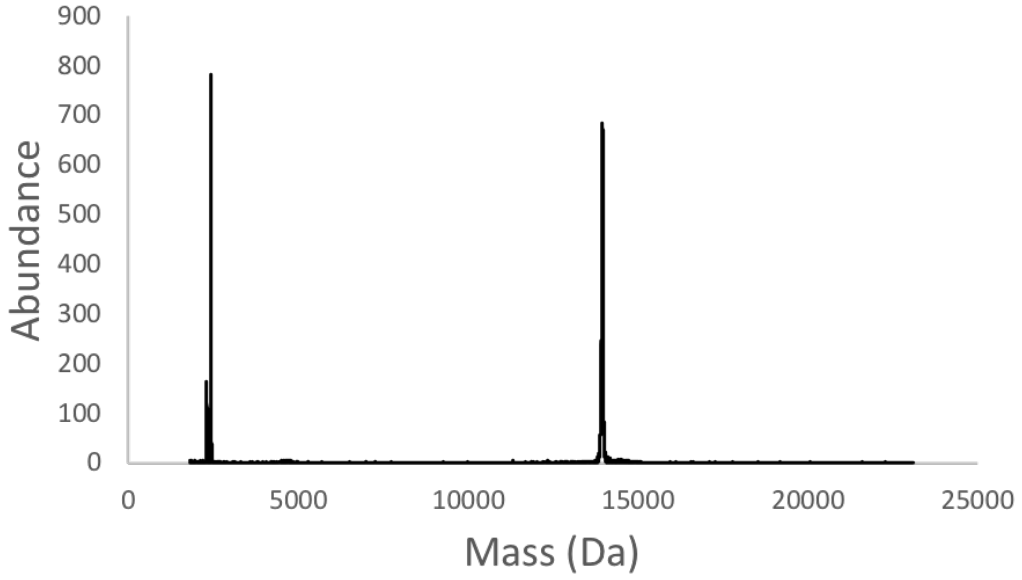

FIGURE 6.3: MALDI- TOF of a single peak taken from the SEC of PduA and L20. Peaks at ~2000 and ~14,000 correspond to P18 and PduA respectively, indicating that both L20 and PduA are present in the sample and the two polypeptides have eluted at the same point, indicative of an interaction between the two.

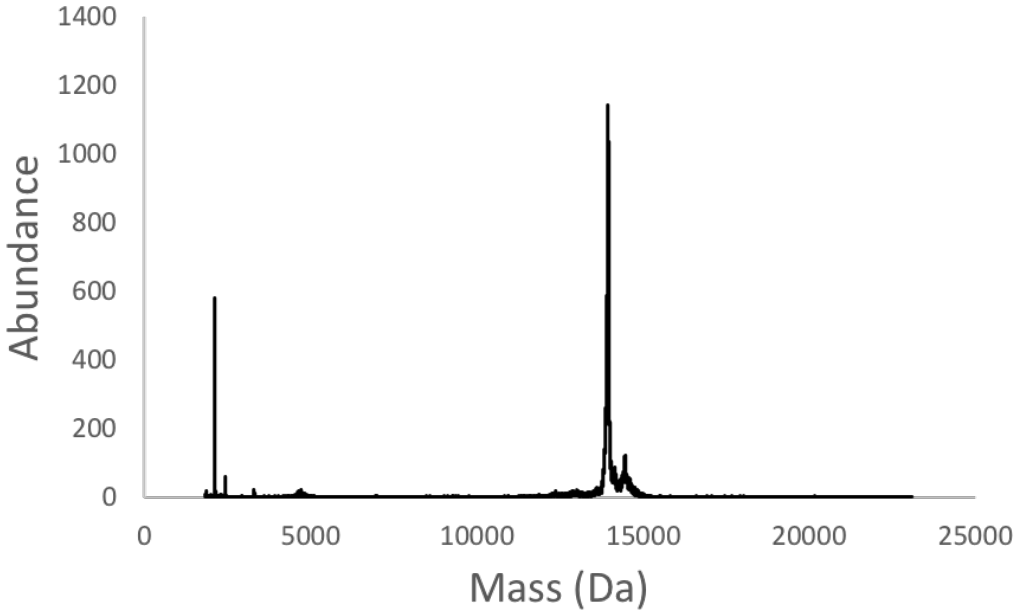

FIGURE 6.4: MALDI- TOF of a single peak taken from the SEC of PduA and D18. Peaks at ~1800 and ~14,000 correspond to P18 and PduA respectively, indicating that both D18 and PduA are present in the sample and the two polypeptides have eluted at the same point, indicative of an interaction between the two.

6.2.2 PduA induces a conformational change in the secondary structure of L20.

The CD spectra of the known Pdu targeting peptides (section: 5.3 figure: 5.2) has shown that the D18 and P18 peptides are alpha helical in their conformation. However, L20 is largely unstructured in solution. It is widely understood that unstructured regions of proteins often undergo conformational changes following complex formation with their partner. Therefore, L20 was added to a sample of PduA in order to observe any changes in helicity.

Following the addition of L20 to 1 mg/ml PduA (in equimolar concentration) there was a shift in the spectra (figure: 6.5). Notably an increase around 222 nm occurred, indicating an increase in the total percentage of α- helices present. From this it can be inferred that either L20 becomes at least partially helical following binding to PduA, or the addition of PduL infers a conformational change to PduA, which promotes α-helix formation.

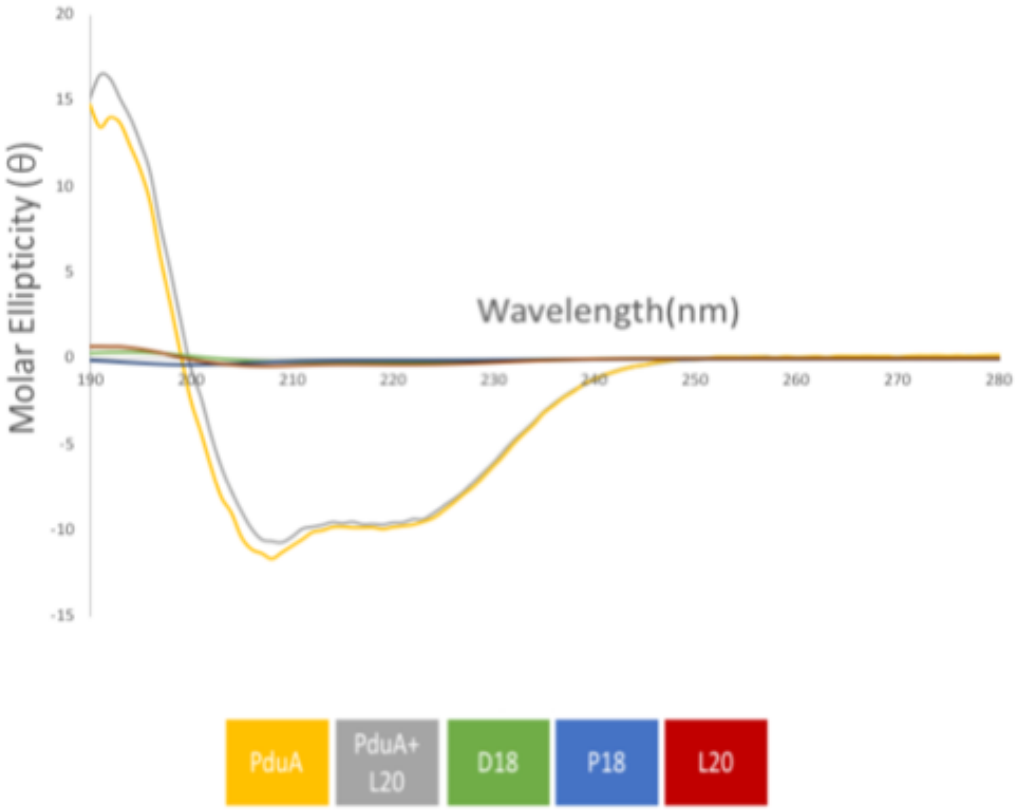

FIGURE 6.5: CD spectra of PduA (yellow) and PduA and L20 (grey) a notable increase in the 222 nm peak is observable following the addition of the L20 peptide, indicative of an increase in the overall proportion of alpha helices, and therefore indicative of an interaction.

6.2.3 The thermal stability of PduL is altered by the addition of PduA.

Assessing the thermal stability of proteins is a key method of probing their interactions. Here, a thermal shift assay was employed to assess the interactions between PduL and the hexameric and trimeric shell proteins: PduA and PduB from *L. reuterii* (figure: 6.6). PduB from *L. reuterii* is homologous to PduB' from *C. freundii*. PduB' has previously been shown to not interact with internalised enzymes [23], so this makes a useful control. Monitoring the fluorescence of SYPRO orange as temperature increases produces a melt curve, where fluorescence peaks as the protein unfolds. This is due to the hydrophobic core of the protein becoming exposed and binding to the dye.

Following the addition of PduL to PduA and PduB, a clear shift in the melting point of PduL is observable. Here, where PduL present with PduA the melting temperature is higher compared to PduL alone and where PduB is added. This increase in thermal stability is indicative of an interaction between PduA and PduL, where the substantially more stable PduA binds to PduL, resulting in an increase in overall thermal stability. As anticipated, no shift in thermal stability is observable where PduL is added to PduB.

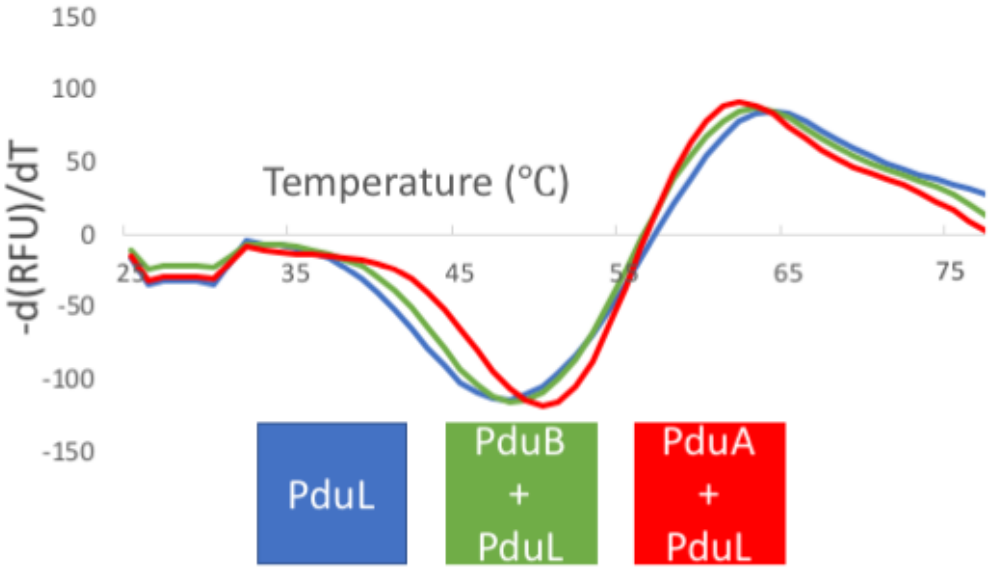

FIGURE 6.6: Average melting profiles of PduL (blue) and PduL in the presence of PduA (red) and PduB (green). A clear shift is notable where PduL is added to PduA, however not where it is added to PduB. This indicates an interaction between PduL and PduA but not PduB.

6.2.4 Addition of PduL to PduA elicits a change in migration when assessed with Native PAGE.

In the previous chapter Native PAGE was highlighted as a powerful yet simple tool for analysing the oligomeric states of proteins and protein complexes. Here, that technique was employed again to assess the interaction between PduL and PduA and the apparent lack of interaction between PduL and PduB.

The migration of PduA, PduB and PduL were assessed using a 4- 20 % gradient Native PAGE. (figure: 6.7). It can be seen that PduL and PduB migrate as two distinct bands. This is in accordance with the previous result that the melting point of PduL is unaltered in the presence of PduB (figure: 6.6). Therefore, it is unlikely that the two proteins interact. It is difficult to assess whether the migration of PduA is altered, as both PduA and PduL migrate similar distances on a 4- 20 % gradient gel. To counter this a fixed 6 % gel was run with increasing concentrations of PduL titrated into PduA in molar ratios from 0:1-3:1 (figure: 6.8). This enables the assessment of the effect of PduL on PduA migration and also provides an estimate of the stoichiometry of the interaction.

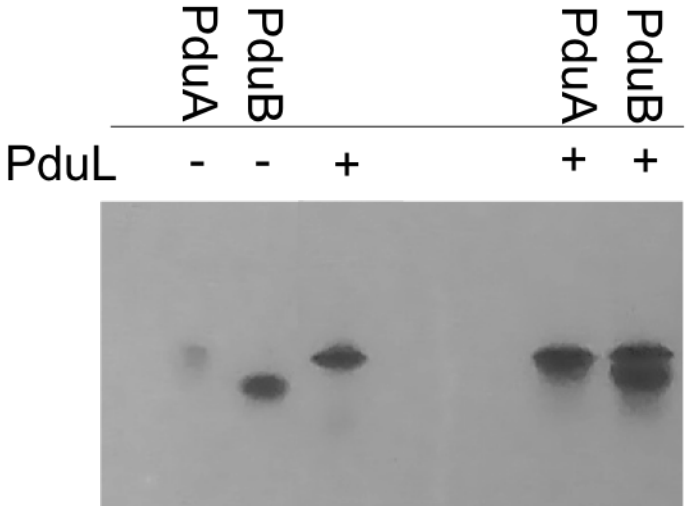

FIGURE 6.7: 4-20 % Native PAGE of PduA, PduB and PduL as well as PduA and B in the presence of PduL. Despite all migrating to similar positions, PduB and L can be seen as two individual bands. It is difficult to determine whether the migration of PduA and PduL has shifted following their combining, running as a single band, or are running as two distinct bands which overlap.

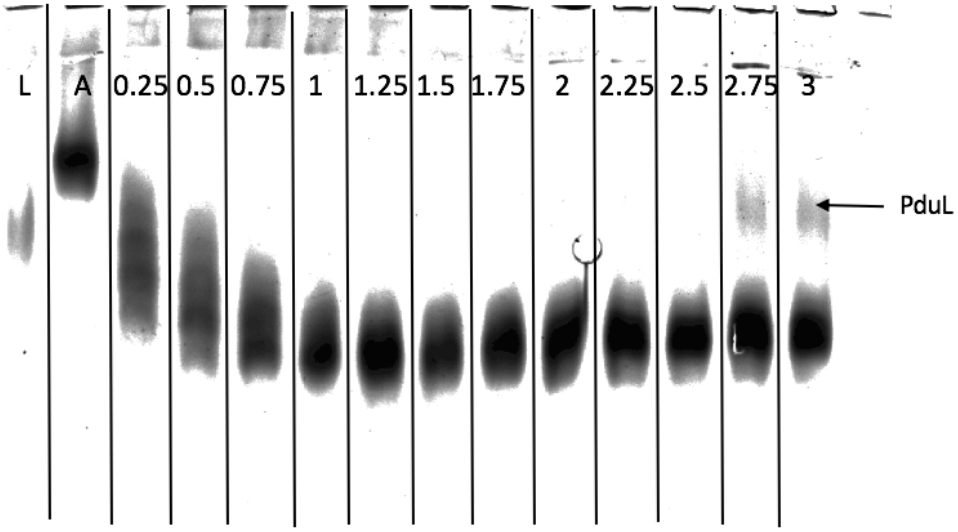

FIGURE 6.8: 6 % fixed Native PAGE of PduA with increasing proportions of PduL (0:1- 3:1 molar ratio of L:A). A clear shift in the mobility is observable with the complex traveling further than either PduA or PduL. Strangely, before total saturation of PduL there are not three distinct bands representing the complex and free proteins, but instead a gradual smear before a single band is observable at a 1:1 ratio. At a molar ratio of 3 PduL's: 1 PduA the band corresponding to free PduL re- emerges, indicating total saturation.

The use of a 6 % fixed gel was sufficient to show a difference in the migration of PduA compared to PduL (figure: 6.8). What is more unexpected is that following addition of PduL the migration of the suspected complex increases, running further than either PduA or PduL individually. Additionally, several poorly distinguished bands are present before total saturation of PduA. These bands are visible at PduL: PduA ratios of 0.25-1:1, possibly corresponding to partially saturated species. PduA appears to be fully saturated at a ratio of 1:1 PduL dimers to PduA hexamers and the re- emergence of PduL is visible at a ratio of around 2:1.

To determine whether the increase in mobility definitely reflected a change in oligomeric state of PduA, a western blot was performed using an anti- his antibody (figure: 6.9) PduL had it's his- tag cleaved and therefore, would not stain. Here, two identical Native PAGE gels were run. A western blot was performed on one (section: 2.5.3) and the other was stained using Coomassie stain (section: 2.4.1). Both PduL and PduA are clearly visible on the Native gel. However, only PduA can be seen on the

western blot. The change in PduA mobility is identical when assessed with both Native PAGE and western blot. This demonstrates that PduA does become more mobile following addition of PduL and it is not an increasing concentration of PduL which interferes with migration.

One explanation for this unexpected result is that it is the change in net charge of the complex which results in the increased migration, rather than a decrease in mass. PduL has a lower pI than PduA and so at pH 8.5 (running buffer of Native PAGE) PduL would be substantially more negative. Additionally, complex formation may reduce the solvent accessibility of positively charged residues. These two factors may be sufficient to lower the net negative charge of the complex, facilitating increased migration.

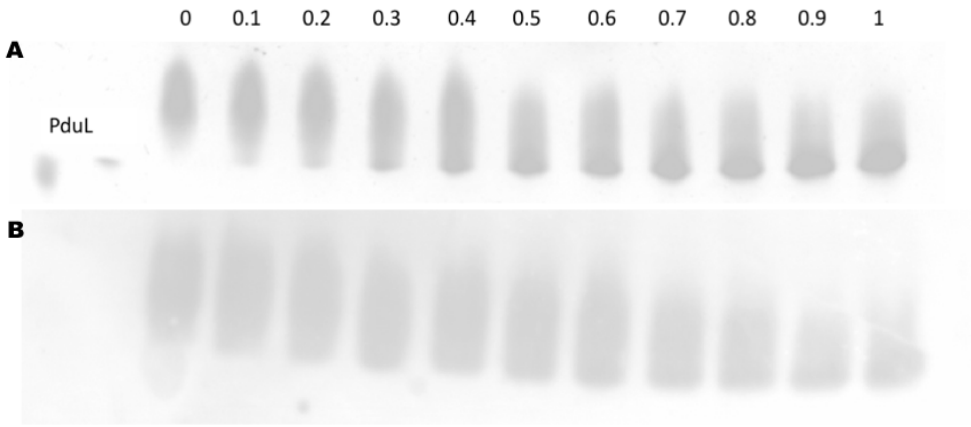

FIGURE 6.9: Native PAGE (A) and the corresponding Anti- His Western blot (B) of PduA with increasing molar ratios of PduL (0:1- 1:1) to PduA. The increasing mobility of the band can be attributed to PduA as here PduL has had its His- tag cleaved and therefore is not stained (as evidenced by the absence of a PduL band in B but not A).

6.2.5 Native PAGE reveals four intermediate states are present before total saturation of PduA.

Before total saturation of PduA (figure: 6.8) several individual bands can be seen, although they are not sufficiently resolved to count the individual bands. Therefore, a repeat of this titration experiment using a 4- 20 % gradient gel was performed (figure: 6.10). Here, an extended runtime (3 hours vs. 2) allowed for better resolution.

A clear shift in the migration of PduA is observed with four distinct bands being present before total saturation of PduA (figure: 6.10). Total saturation occurs at a ratio of 1:1 PduA: PduL. The identity of the bands prominently seen were not immediately obvious. However, presumably the top band corresponds to PduA with no PduL bound and the bottom band corresponds to totally saturated PduA. The two intermediate bands must therefore be partially saturated species. The apparent ratio of 1:1 presents issues, as a single PduA hexamer binding to a single PduL dimer would not yield any partially saturated intermediates. Instead, this interaction would be binary whereby, PduL is either bound or not. However, as discussed in depth in chapter: 3, PduA is unlikely to be a single hexamer in solution or when assessed using Native PAGE. An oligomer of PduA hexamers binding to several PduLs could produce the multiple bands seen here (figure: 6.10).

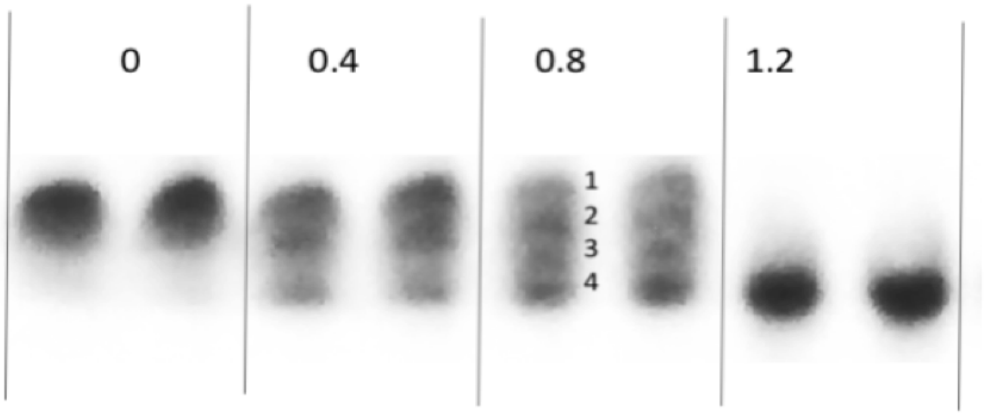

FIGURE 6.10: 4-20% gradient gel of PduA with increasing molar ratios of PduL (0-1.2:1). At a ratio of 0.8 PduL: 1 PduA, four bands can be seen between free (represented by the uppermost band) and totally saturated (represented by the bottom band) PduA.

The question remains as too whether PduL is binding to the surface of the PduA hexamers or at the interface between hexamers. If PduL binds to the surface of PduA hexamers, then an interaction with the non- tessellating mutants previously described (K26D and PduA$_{GST}$) should be easily observable.

6.3 PduL does not interact with non- tessellating variants of PduA.

Chapter: 3 presented evidence that PduA is not present as single hexamers in solution, or when assessed with native PAGE, but rather multimers thereof. Therefore, the interactions between PduL and tessellating and non- tessellating PduA were investigated. There are several factors which promote a model of binding where PduL binds to a site spanning multiple PduA hexamers, including:

1) The data presented in figure: 6.10 can not be reconciled with a model where A single PduA hexamer binds a single PduL dimer.

2) The previous model, suggesting that signalling peptides bind to the C- terminus of PduA [51, 52] is not compatible with current thought regarding the orientation of PduA [31, 30] (in respect to the whole microcompartment).

The non- tessellating PduA mutants discussed in chapter: 3 are an excellent avenue for exploring this possibility; if the PduL binding site does span multiple hexamers, no interaction should be detectable between these mutants and PduL.

6.3.1 Non- tessellating mutants of PduA do not interact with PduL as assessed by Native PAGE.

Here, (figure: 6.11) a 4-20 % Native PAGE was run with PduA and both K26D and Pdu_{GST}. As seen previously, the addition of PduL to PduA produces an increase in mobility, with both proteins running as a single band. In the case of both K26D and $PduA_{GST}$ no effect on mobility was observed and the original band corresponding to PduL persisted. This would suggest that tessellation of PduA is necessary for the binding of PduL and therefore, the binding site is very likely at the interface between two PduA hexamers.

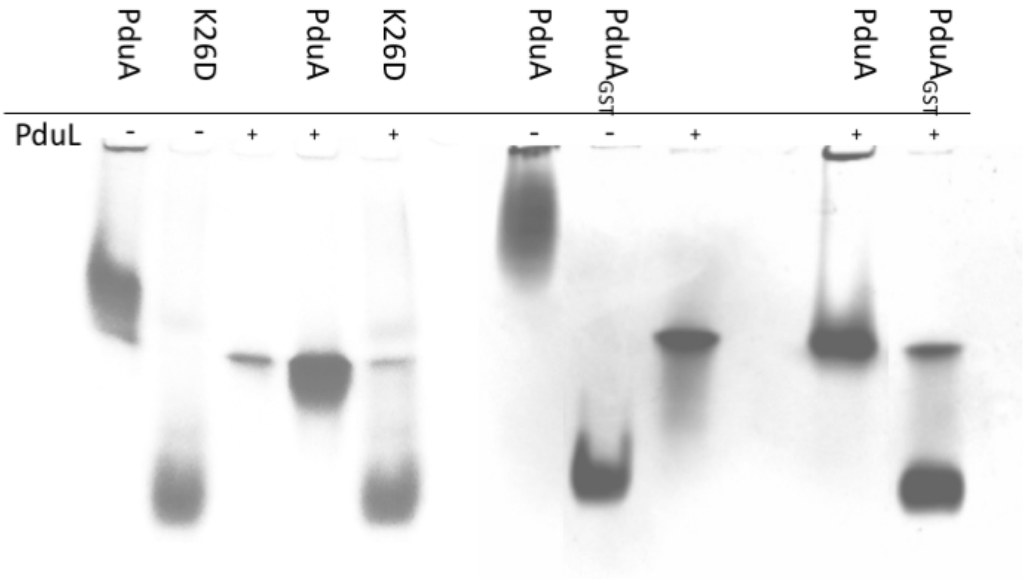

FIGURE 6.11: Native PAGE with PduA and non- tessellating mutants K26D and PduA$_{GS}$T. The presence or absence of PduL is indicated by a + or - respectively. The addition of PduL produces a substantial shift in the position of PduA. Both non- tessellating variants remain in their original positions following addition of PduL and the band corresponding to PduL persists.

One anecdotal observation is that the age of a PduA sample greatly influences its mobility on Native PAGE. Samples several days old will migrate significantly less than fresh samples following SEC. PduA oligomerises over time to become sheets and tubes. These large oligomers are presumably responsible for the reduction in PduA mobility and the increased propensity to smear over time. Remarkably, this oligomerisation appears to be reversed by the addition of PduL. The complex consistently runs to the same position regardless of the initial starting point of PduA. This indicates that whilst the initial sample of PduA might vary in molecular weight, the complex is always the same size.

6.3.2 Non- tessellating mutants of PduA do not interact with PduL as assessed by Circular Dichroism.

The Native PAGE experiments demonstrated an apparent lack of any interaction between the non- tessellating PduA mutants and PduL. Therefore, the CD experiment from section: 6.2.2 was repeated using PduA$_{GST}$ as oppose to PduA. Here, following

the addition of the L20 peptide to PduA$_{GST}$ there was no change in the CD spectra. Notably, the increase in the 222 nm peak seen in figure: 6.5 was absent, indicating no overall change in the helicity of either PduA or the L20 peptide. This result supports the previous Native PAGE results, and cements the hypothesis that tessellation of PduA is critical for the interaction with the PduL N- terminus.

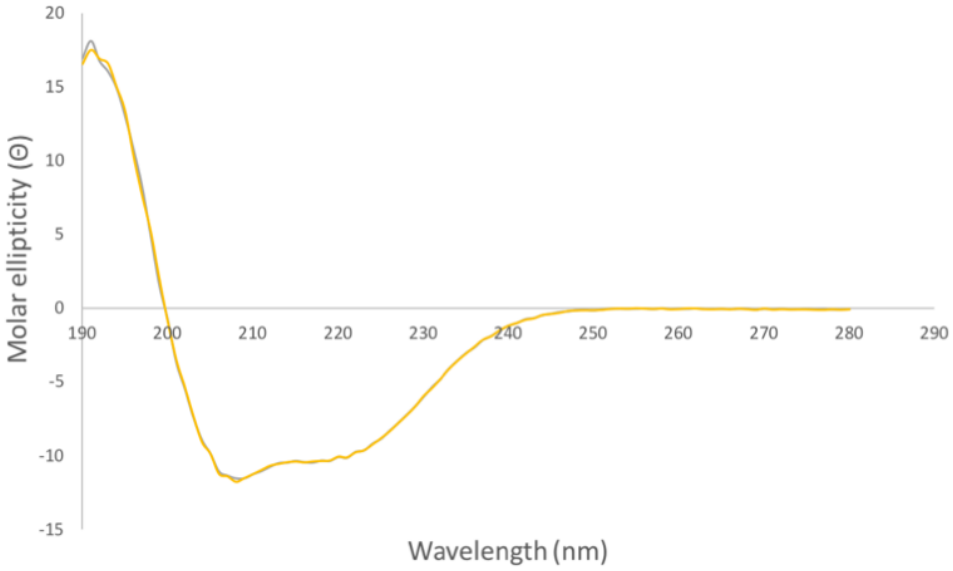

FIGURE 6.12: CD spectra of PduA$_{GST}$ (yellow) and PduA$_{GST}$ with the addition of an equimolar amount of the L20 peptide (grey). The two spectra are near identical, with no change in the 222nm peak as seen where PduL is added to PduA, thus demonstrating there is no change in the proportion of alpha helices present, and therefore no interaction between PduA$_{GST}$ and L20 can be inferred.

6.4 Addition of PduL inhibits the formation of PduA nanotubes.

PduA nanotubes are known, large oligomers of PduA. If there is some effect of binding PduL on the overall oligomeric state, and consequently the molecular weight of the complex, then there should be some morphological change apparent when PduL is added to PduA nanotubes. Typically, PduA nanotubes are assembled by concentrating PduA to 5 mg/ml in low (50 mM) NaCl concentrations. These conditions promote the rapid multimerisation of PduA into nanotubes. When observed under

TEM these nanotubes have a diameter of 20 nm and lengths varying from 400- 1000 nm. Additionally, the tubes tend to line up in parallel to one another, co- localising into ordered aggregates on the grid (figure: 6.13A).

In contrast when PduA tubes were prepared in the manner described above, with the addition of PduL a staggering difference in tube morphology was observable (figure: 6.13B). Nanotubes were notably shorter (~150 nm) and what appears to be rings of PduA were visible. Additionally, the tubes no longer lined up in the ordered manner seen above but instead co- localised in a more chaotic state, pointing in various orientations with respect to one another. However, the average diameter of the tubes remained unchanged.

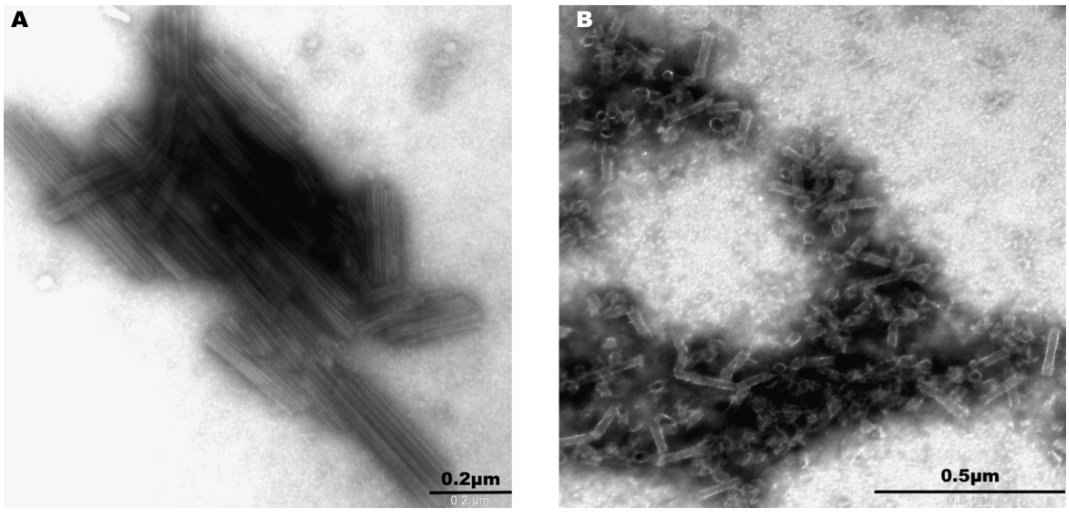

FIGURE 6.13: TEM image of PduA nanotubes formed at a PduA concentration of 5mg/ml and NaCl concentration of 50 mM (A) . The long, straight, ordered tubes, with a diameter of 20 nm are typical of the tubes previously reported to be formed by PduA. TEM image of PduA nanotubes formed at a PduA concentration of 5 mg/ml and NaCl concentration of 50 mM with an equimolar amount of PduL present (B). These tubes maintain the typical diameter of 20 nm, however they are far shorter than tubes formed in the absence of PduL, and are arranged in a far more disordered manner.

As described in chapter: 1, section: 1.3 PduA like proteins are involved in the formation of both flat facets and curved edges of BMCs. Therefore, PduA like proteins are stable in either a planar or curved assembly. In PduA tubes there must be an angle present at the hexamer- hexamer interface to allow for the curvature of the tube. It is plausible that the peptide inserts itself at the interface between hexamers.

This insertion would distort the interface, eliminating the curvature and reducing the formation of tubes. This explanation would account for both the data presented here (figure: 6.13) and also the observation that regardless of the initial size of a PduA oligomer, the complex always runs to the same position on a Native PAGE (figures: 6.7, 6.8, 6.10 and 6.11). If the PduA initially present in a Native PAGE is in a curved conformation as some kind of "prototube", then the addition of PduL would distort these angled interfaces, potentially straightening them or promoting disassembly.

Shown in figure: 6.14 is a comparison between the curved and flat interface between hexamers from PDB: 5B74. A prospective binding site located at the interface is visible where hexamers interact 180 ° relative to another vs 150 °. The binding of PduL could potentially inhibit the curved conformation by inserting itself in the cleft and forcing the adoption of the flat conformation, thus inhibiting tube formation.

A model where the peptide binds at the interface between two hexamers also accounts for the difference seen between WT and "empty" BMCs. Empty BMCs, produced by bacteria which only express the shell proteins are notably smaller than WT BMCs (~60 vs. ~120 nm) [95]. This difference in size would be produced where signalling peptides bind to the interface and promote the formation of the flat surfaces rather than curved ones. Consequently, where there are larger flat facets which are incorporated into the BMC the entire BMC must be larger.

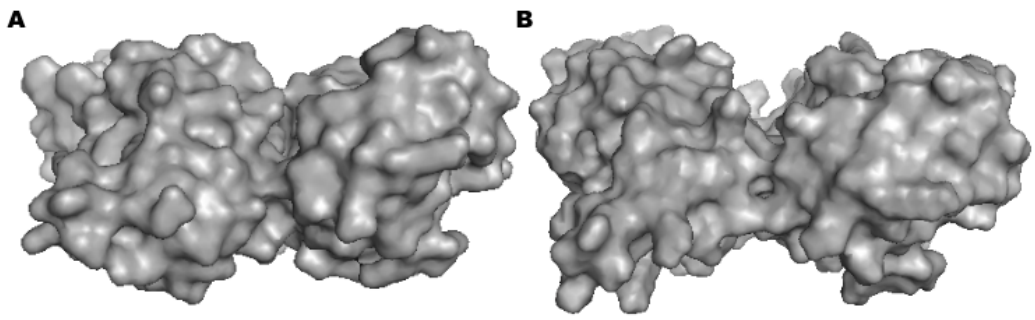

FIGURE 6.14: Comparison of the interface between hexamers found in the crystal structure PDB: 5B74. Where hexamers form the curved edge of the shell a cleft at the interface is occluded (A). Where hexamers form the flat facets of the compartment (B) this cleft is exposed.

6.5 PduK interacts with PduL.

Following the investigations into the interaction between PduA and PduL, PduK was assessed to determine if it interacted with PduL. It has previously been shown that PduK interacts with PduP [29] which features a very similar N– terminal extension to PduL, and so it seemed likely the two would form a complex.

Native PAGE was performed with PduK and PduL (figure: 6.15). Interestingly, PduK does not run as a single band but instead two distinct bands, despite being only visible as a single band on SDS- PAGE. The higher band runs to the a similar position as PduA in respect to PduL (figure: 6.7) and has a similar smeary appearance. Presumably as with PduA, this band corresponds to an assembly of multiple hexamers and the heterogeneity of these assemblies produces the smeared appearance. The lower band however, is far more well defined and runs to a similar position as the non-tessellating mutant of PduA ($PduA_{GST}$). This may be indicative of PduK existing at an equilibrium between hexamers, and multimers of hexamers in solution. The shift in migration seen where PduL is added to PduA is less apparent where PduL is added to PduK. However, the upper smeared band appears more well defined, with a significant increase in intensity. Therefore, whilst the addition of PduL appears to induce a change in the appearance of PduK on Native PAGE, this is not sufficient to conclude a complex has formed. Consequently, other techniques were employed.

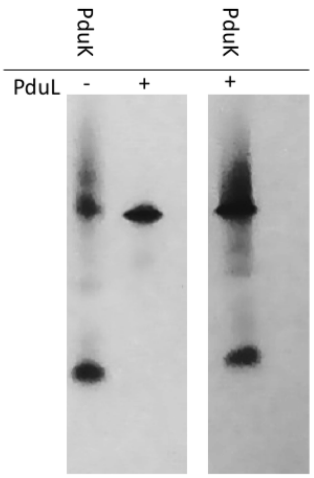

FIGURE 6.15: Native PAGE of PduK and PduL. PduK features two bands, one at a similar position to where PduA typically migrates which likely represents the previously described tessellated hexamers. The lower band is therefore likely a single hexamer, as seen when analysing the non- tessellating PduA mutants. Following the addition of PduL to PduK a change can be seen in the upper, but not lower band.

6.5.1 Using fluorescence quenching to measure the interaction between PduL and PduK.

Unlike PduA, PduK features a tryptophan residue at position 42, endowing it with an intrinsic fluorescence at 350 nm. This tryptophan has been previously exploited to characterise the interaction between PduK and PduP (using only the 18 N- terminal residues of PduP) [29]. Here, full length PduL was titrated into PduK and the fluorescence was monitored. The advantage of performing this experiment with PduL is that PduL features no residues with intrinsic fluorescence. Consequently, the full protein may be used rather than just the encapsulation peptide.

Following titration of PduL into PduK a loss of the fluorescence signal occurred rapidly before total saturation at a ratio of around 1:1 (figure: 6.16), consistent with results for the interaction between PduA and PduL. A value for the total concentration of the complex was calculated from the fluorescence values and a hyperbolic

curve was fitted to the data using the formula:

$$y = m * xk + x$$

The Kd was calculated from this curve to be 130.46 nM. The results following Native PAGE (section: 6.2.3, figure: 6.7) were uncertain as to whether PduK and PduL interacted. However, this experiment was able to definitively demonstrate that PduK interacts with PduL.

Full length PduK was shown in chapter: 3, section: 3.6 to form flat sheets whereas, the truncated A100>STOP PduK formed tubes and "pseudocompartments". This phenomena provided an opportunity to investigate the preferential binding of PduL to shell proteins in a planar arrangement (discussed in section: 6.4). If PduL binds at the exposed interface present between two flat hexamers, then the affinity of PduL for truncated A100>STOP PduK should be reduced.

PduL was titrated into PduK A100>STOP and the quenching experiment repeated (figure: 6.17). The Kd was calculated to be substantially lower, at 853.15 nM. Whilst this change was anticipated, there are several factors which may play in to this change in affinity. The most obvious is that the C- terminus of PduK interacts directly with PduL. However, this is unlikely as sequence alignment shows the overall fold of PduK is the same as PduA (figure: 3.16). Therefore, it is extremely likely that the C- terminus is present on the exterior of the BMC, meaning that an interaction with the internalised PduL would be impossible. Additionally, the Tryptophan at position 42 which is quenched is present on the opposite surface to the C- termini, making it unlikely that this is the surface to which PduL is binding.

The most plausible explanation for the increase in Kd (figure: 6.17), is that the increased curvature seen in the morphology of A100>STOP macrostructures reduces access to the proposed binding site (figure: 6.14). This explanation is consistent with section: 6.4 whereby, addition of PduL to PduA interfered with the formation of nanotubes. This evidence, combined with the fact that non- tessellating mutants of PduA are not able to interact with PduL (section: 6.3) points directly to a model where signalling peptides bind shell proteins at the interface between shell proteins.

No other explanation is able to satisfy all experimental observations, whilst also being reconcilable with the understanding that the C- termini of hexameric shell proteins are present on the exterior of BMCs

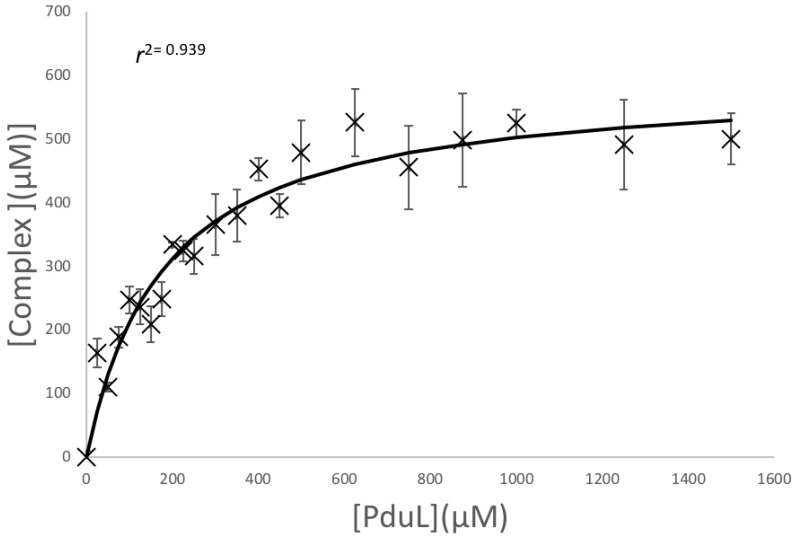

FIGURE 6.16: Plot of concentration of PduL against total concentration of the protein complex, as calculated from the flourescence values. A clear saturation curve is seen and a trend- line fitted using a hyperbolic function, a kd of 130.46 nM was calculated from this curve.

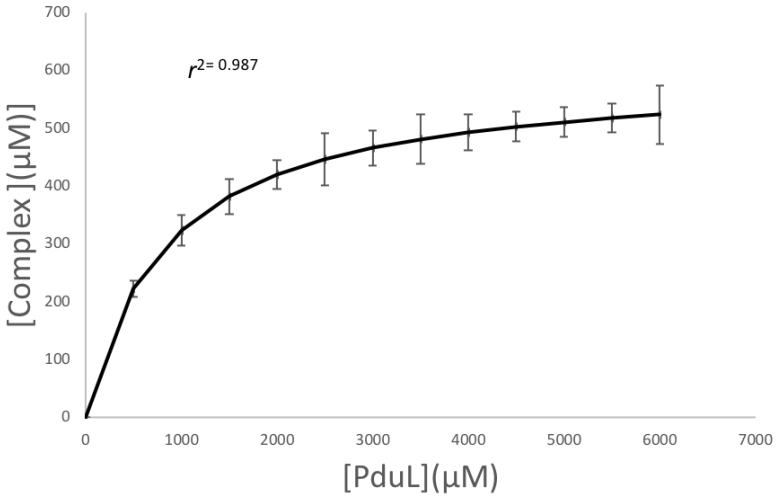

FIGURE 6.17: Plot of concentration of PduL against total concentration of the protein complex, as calculated from the flourescence values. A clear saturation curve is seen and a trend- line fitted using a hyperbolic function, a kd of 853.15 nM was calculated from this curve.

6.6 PduA facilitates optimised substrate transfer between PduP and PduL.

The enzymes PduP and PduL act in a synergistic manner within the Pdu BMC. PduP acts to convert Propionaldehyde into Propionyl- CoA, reducing NAD+ into NADH in the process. Propionyl- CoA is then converted into propionyl phosphate and CoA by PduL in a process which uses a single phosphate ion. The regenerated CoA can then be used in the processing of another propionaldehyde by PduP, negating the need to transport a large molecule across the BMC shell. Previous work has demonstrated that targeting of non- native enzyme pairs into the BMC lumen results in greater yields of final product (when compared to those same enzymes in solution [54, 55]). This increase in yield is hypothesised to primarily result from the localised increase in substrate concentration, where diffusion out of the BMC is limited. However, there may also be an increased rate of substrate transfer between enzymes as a product of them being effectively tethered together on the BMC shell. This would reduce the diffusion distances of substrates and products between enzymes and facilitate substrate transfer by approximation.

Two factors described previously present an opportunity to investigate substrate transfer by approximation:

1) The discovery that PduA does not primarily exist as single hexamers in solution but rather multimers thereof (Chapter: 3)

2) These multimers are capable of binding to the encapsulated enzymes (Section: 6.2).

As these multimers do not form closed structures, any change in substrate transfer should entirely be a product of increased proximity between enzymes.

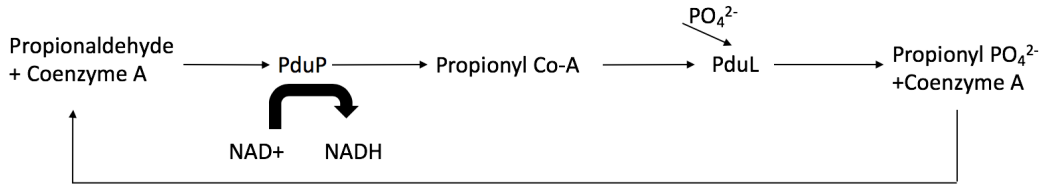

FIGURE 6.18: A Diagramatic representation of the substrates and products of PduP and PduL, demonstrating CoA is recycled by PduL and used by PduP, the only molecules required to enter the system are propionaldehyde and phosphate, as NAD+ is recycled by PduQ.

6.6.1 Purification and preparation of PduP.

PduP from C. freundii was expressed using a pET14b vector in *E. coli* and produced in the method outlined in chapter 2 (Section: 2.3.1). Analysis with SDS- PAGE (figure: 6.19) revealed a single band at 50 kDa following Nickel IMAC (consistent with the proteins theoretical weight of 48.6 kDa). PduP eluted at 15.2 ml following size SEC on an S200 column, a result consistent with PduP being a monomer. As with PduL, cleavage of the his- tag to fully expose the N- terminal signalling sequence was desirable. This was achieved through overnight incubation with thrombin at 4 °C. The larger size of PduP makes it more difficult to determine whether cleavage of the tag was successful using SDS- PAGE. However, the use of reverse IMAC following incubation with thrombin should ensure all proteins with their tag intact are retained in the column. Following all purification steps PduP was produced to a high homogeneity, suitable for all subsequent experiments.

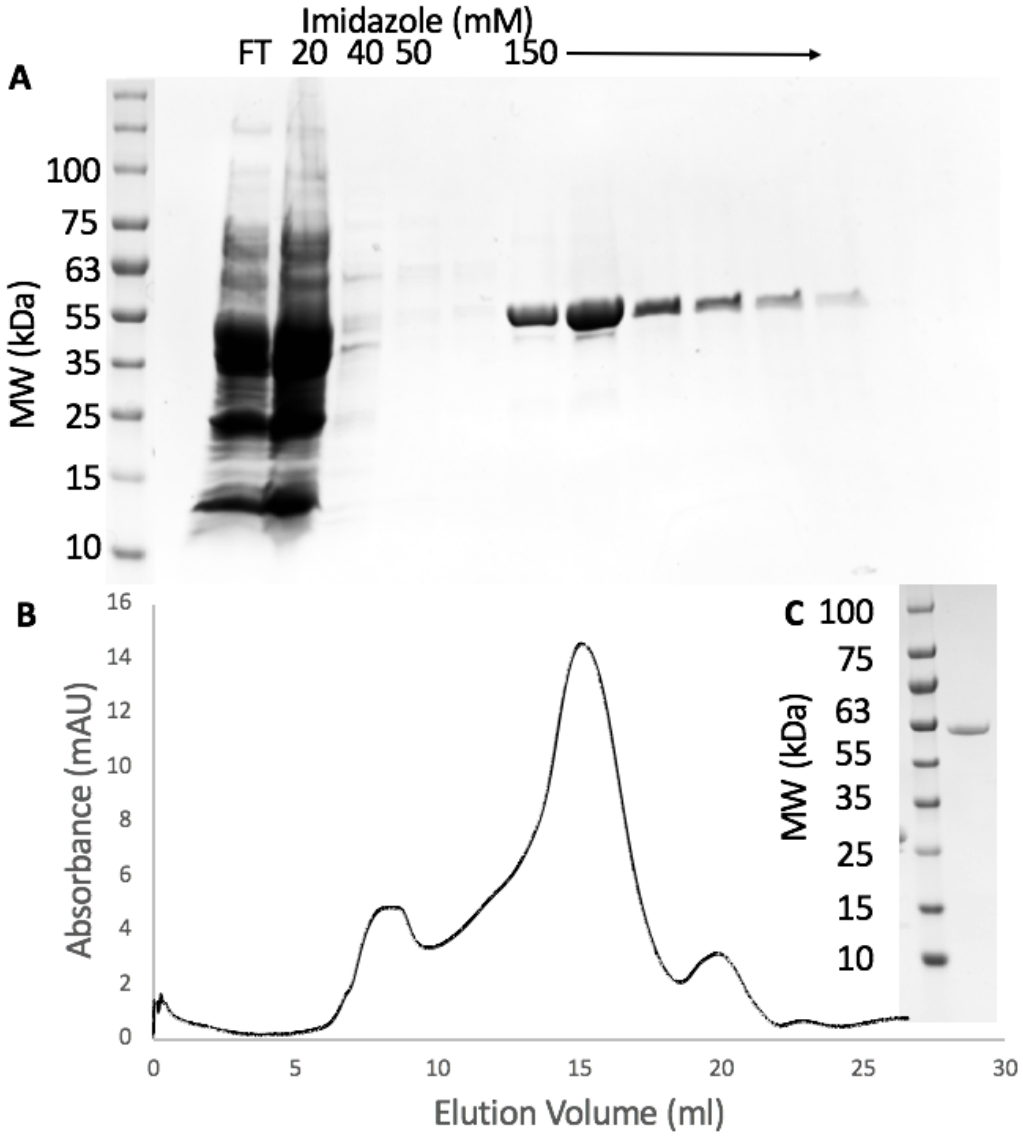

FIGURE 6.19: Purification of PduP. SDS- PAGE of IMAC fractions (A). PduP elutes following the addition of 80 mM imidazole at a mass of ~50 kDa, consistent with the theoretical mass. SEC of the purified sample using an S200 column (B), yielded a major peak at 15 ml, consistent with the mass of a monomer (~50 kDa). The SEC profile showed several peaks, despite the SDS- PAGE appearing relatively homogenous. This is likely a product of PduPs tendency to aggregate. Therefore, whilst only a single protein is present, multiple different oligomers corresponding to multiple sizes are present. SDS-PAGE of the purified sample (C), showed the protein was purified to a high homogeneity.

6.6.2 Substrate transfer between PduP and PduL.

PduP and PduL were added to a reaction mixture as described in section: 2.7.2 in a 1 ml cuvette. The reaction was initiated by the addition of Propionyl- CoA. The reaction was monitored by measuring the absorbance at 340 nm over time, as NADH will absorb UV at 340 nm whereas, NAD+ will not. By adding Propionyl- CoA, PduP is unable to reduce NAD+ until free CoA is made available from Propionyl-CoA by PduL. Therefore, the rate of reduction of NAD+ corresponds to the conversion and transfer of Propionyl- CoA by PduL. Following this experiment the same was repeated, with the addition of PduA. Where PduA was added the reduction of NAD+ was completed after 300 seconds and a specific activity can be calculated to 1.5 $\mu g min^{-1} mg^{-1}$. Whereas, where no PduA is present the reaction completes after 1400 seconds (figure: 6.20), with a specific activity of 0.83 $\mu g min^{-1} mg^{-1}$. As PduA has no intrinsic enzymatic activity, this would seemingly imply that; the presence of PduA increases the transfer of CoA from PduL to PduP based on proximity effects alone. However, other options may also account for this apparent change in turnover. Notably, the notion that the binding of PduL or PduP to PduA imparts some conformational change on the enzyme, which in turn has an effect on the activity of either enzyme.

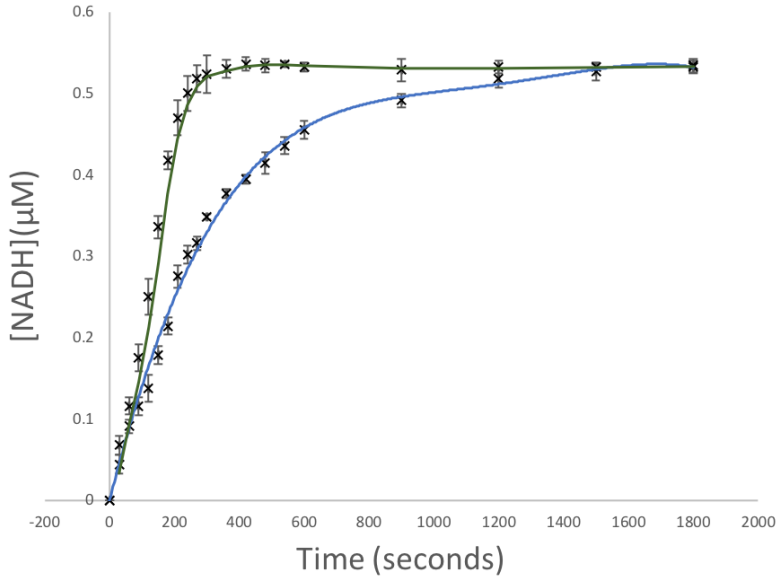

FIGURE 6.20: Concentration of NADH plotted against time. The concentration of NADH was calculated using an extinction coefficient of 6220. Shown on the graph are the changes in NADH concentration with PduP and PduL alone (trend line coloured blue) and PduP and PduL in the presence of PduA (trend line coloured green). The sharp increase in rate is evident when comparing the two curves, and specific activities of 1.5 $\mu g min^{-1} mg^{-1}$ and 0.83 $\mu g min^{-1} mg^{-1}$ can be calculated from the curves for PduA, L and P and PduP and L respectively.

To ascertain whether the binding of PduA effected the activity of PduP, PduP was added alone to a reaction mixture. This mixture was identical to the one used for the previous experiment, although the 100 mM Propionyl- CoA was substituted for 100 mM CoA (as there is no PduL present to catalyse the conversion of Propionyl- CoA into CoA). The absorbance at 340 nm was recorded until the maxima absorbance was achieved (figure: 6.21). This was then repeated with the addition of PduA (figure: 6.22). The specific activity of PduP in the presence and absence of PduA was not significantly different 1.06 vs. 1.12 $\mu g min^{-1} mg^{-1}$. Therefore, this experiment was able to confirm that the addition of PduA does not effect the activity of PduP alone.

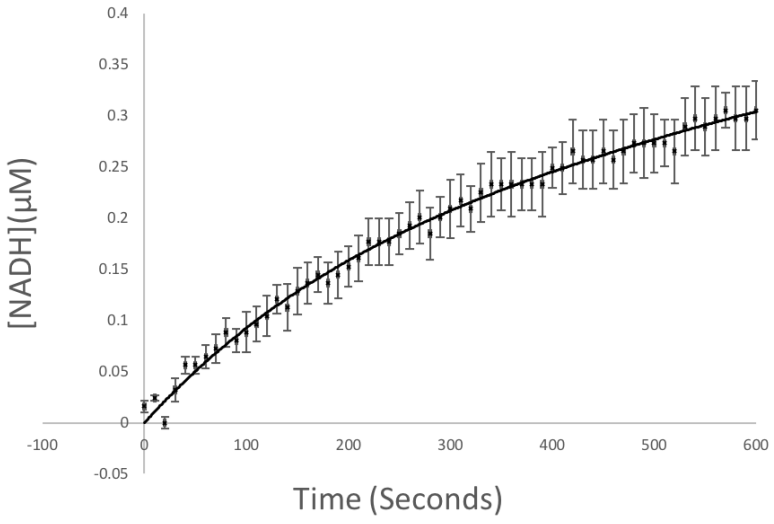

FIGURE 6.21: NADH concentration plotted against time in the presence of 0.2mg PduP. The concentration of NADH was calculated using an extinction coefficient of 6220. The specific activity of this reaction was calculated to be 1.06 $\mu g min^{-1} mg^{-1}$

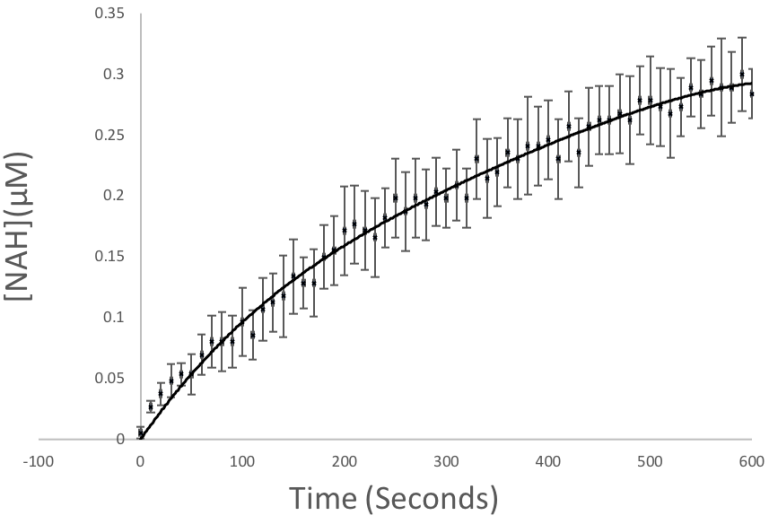

FIGURE 6.22: NADH concentration plotted against time in the presence of 0.2mg PduP and PduA. The concentration of NADH was calculated using an extinction coefficient of 6220. The specific activity of this reaction was calculated to be 1.12 $\mu g min^{-1} mg^{-1}$.

In order to determine whether the addition of PduA influenced the activity of PduL, an alternative assay system was required. None of the substrates or products of PduL feature a unique UV absorbance spectra and so the detection of a species

formed following a secondary reaction was chosen. Ammonium molybdate is commonly used in assays for the determination of inorganic phosphate, as it will react with phosphate in the presence of an acid to form ammonium phosphomolybdate. The reduced compound features a strong blue colour with an absorbance spectra which has a maxima at 832 nm. In this assay the loss of signal at 832 nm is measured and the amount of phosphate can be calculated using a standard curve (Figure: 6.23). As phosphate is used by PduL in a 1:1 ratio with propionyl- CoA in the production of propionyl- phosphate, the amount of phosphate used is equivalent to the propionyl-phosphate produced.

The results of this experiment were unexpected. Samples containing PduL resulted in greater absorbance values than those that did not, indicative of greater levels of phosphate in the sample. It is impossible for the reverse reaction to have produced phosphate as there was no propionyl- phosphate present in the initial sample. Therefore, propionyl- phosphate must be able to react with Ammonium molybdate to form a complex with even greater absorbance at 832 nm. This theory was confirmed following the production of a standard curve for propionyl- phosphate (figure: 6.23). Here, equimolar concentrations of propionyl- phosphate will produce greater absorbance peaks when compared to phosphate. As a result, no meaningful quantitative data was produced using this technique, as it is impossible to discern how much either phosphate or propionyl- phosphate contribute to a peak. It is however, still possible to obtain some rudimentary qualitative data. Where the A_{832} value increases following a reaction, there must be a greater amount of propionyl-phosphate present compared to phosphate.

Figure: 6.24 demonstrated that there is no significant difference in the activity of PduL alone when compared to PduL in the presence of PduA $t(6) = -1.04$, $p = 0.18$. Additionally, substrate transfer from PduP to PduL was detectable where PduP and PduL were added together in the presence of CoA, phosphate and propionaldehyde. Here, PduP must process CoA and propionaldehyde into propionyl- CoA for the PduL reaction to proceed. Therefore, the greater absorbance value detected where PduA was present, indicated a significantly increased substrate transfer between PduP and PduL $t(6) = 11.6$, $p < 0.001$.

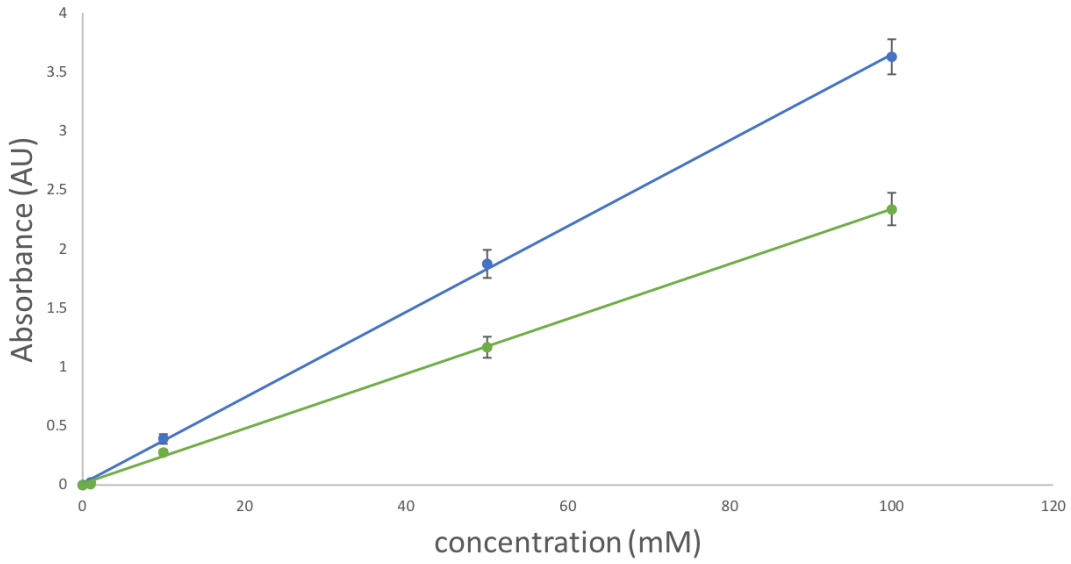

FIGURE 6.23: Phosphate concentration standard curves for both phosphate(green) and propionyl phosphate (blue), the stark difference in the absorbance values obtained from the complex despite equivalent concentrations is evident, with propionyl phosphate producing a far greater signal than phosphate.

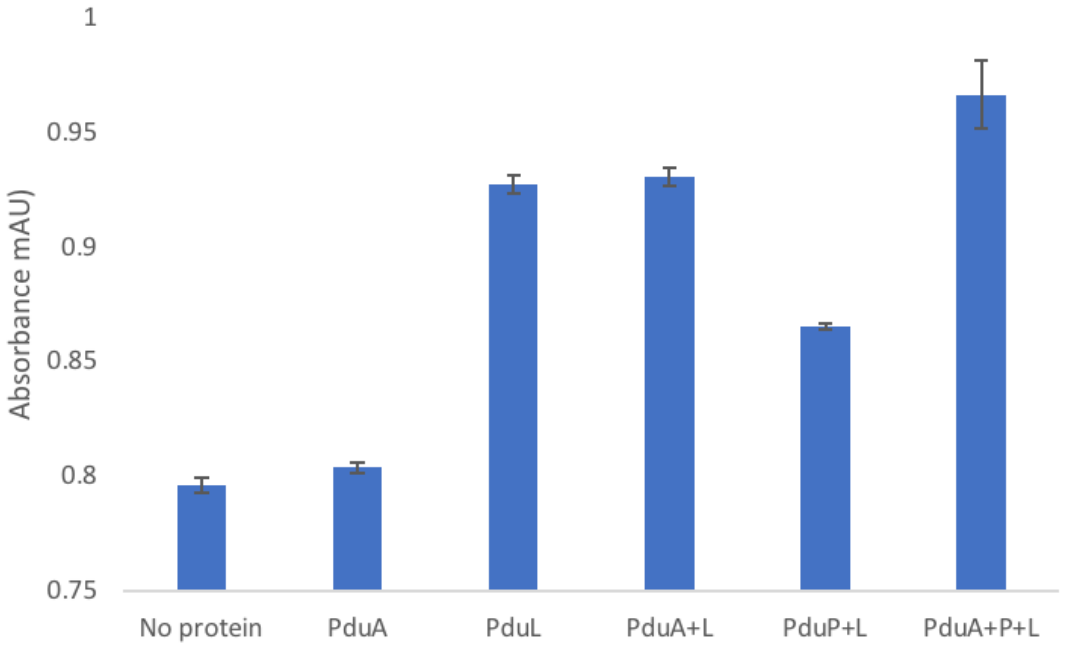

FIGURE 6.24: Bar chart showing A_{832} values following termination of reactions. The increased absorbance seen in samples containing PduL is evidence of enzymatic activity, although the addition of PduA does not effect this. A greater A_{832} value can be seen, however, when comparing PduL and PduP Vs. PduL, PduP and PduA.

6.7 Conclusions and Discussion.

The evidence presented in this chapter sheds new light on the interactions between shell proteins and encapsulated enzymes. A specific interaction between PduA and PduL has been highlighted as an example. Evidence, as suggested by CD and native PAGE shows a clear link between tessellation of PduA and its ability to bind the PduL signalling peptide. This new explanation for the binding of enzymes at the interface between two adjacent hexamers is in agreement with the current thought regarding the orientation of the BMC shell. The PduL peptide binds to what is now thought to be the interior face of the shell. This would permit the sequestration of substrate molecules and volatile intermediates within the BMC lumen, a feat not possible should the associated enzymes be present on the shell exterior.

The binding of enzymes at the hexamer- hexamer interface may also be key to understanding the formation of BMCs. Where shell proteins are expressed in the absence of enzymes, curved structures including tubes and rolled up sheets are commonly observed [25, 17]. However, there must be some planar surfaces present in complete microcompartments, which form the flat facets. Evidence presented here (section: 6.4, section: 6.5.1), demonstrates that addition of PduL to PduA inhibits the formation of nanotubes. Additionally, PduL will bind less readily to PduK where curvature is present in the macrostructures. Taken together, this evidence demonstrates a clear preference for planar shell protein assemblies binding to PduL vs. assemblies where significant curvature is present. Therefore, it is likely that the flat facets of WT BMCs are produced as a result of the helical extension peptides binding at hexamer- hexamer interfaces. This would reduce the permitted curvature between adjacent hexamers as the microcompartment assembles. Consequently, this explanation also accounts for why artificially produced "empty" microcompartments are substantially smaller than their wild type counterparts [18]. Where empty BMCs are synthesised in the absence of enzymes, there is nothing which regulates the curvature formed by shell protein assemblies. This would result in far smaller flat facets, which in turn leads to smaller microcompartments.

The observation that enhanced substrate transfer occurs between PduP and PduL

where PduA is present may prove useful in future industrial applications. Previously, the process of producing enzymes with Pdu signalling peptides and also producing entire BMCs to encapsulate them was complex. However, if enhanced substrate transfer can be achieved through producing two enzymes with the L20 and P18 signalling peptide and adding a single protein, PduA, this would provide a significantly more simple and cost effective method for optimising substrate transfer in industrial applications.

Chapter 7

Modelling the PduA- PduL complex.

7.1 Modelling the PduA: PduL complex.

Substantial data in chapter: 6 has been presented supporting the theory that pdu signalling peptides bind at the interface between hexamers in a flat sheet. Co- crystallisation of PduA and PduL was attempted in order to prove this. However, no crystal containing both PduA and PduL could be obtained. Therefore, an in silico approach was taken.

A pair of hexamers were used as the initial search model, as oppose to previous studies which just used a single hexamer. Unsurprisingly, when a pair of hexamers were used as a receptor in docking simulations the peptide consistently localised to the interface, regardless of which software was used. The CABS- Dock server [81] was particularly useful as it simulates a flexible peptide which folds concurrently throughout docking. This differs from other docking software such as HADDOCK [96] which require input PDB files and therefore allow for no flexibility of the peptide backbone. Additionally, CABS- Dock requires no preliminary data regarding specific residues as inputs, and therefore eliminates any biases which may be introduced through this. The model produced in CABS- Dock was then refined using the Flexpepdock server [84] which produced binding energy scores. This allowed for comparison of the models produced with previous models, and also those with the peptide bound to a single hexamer.

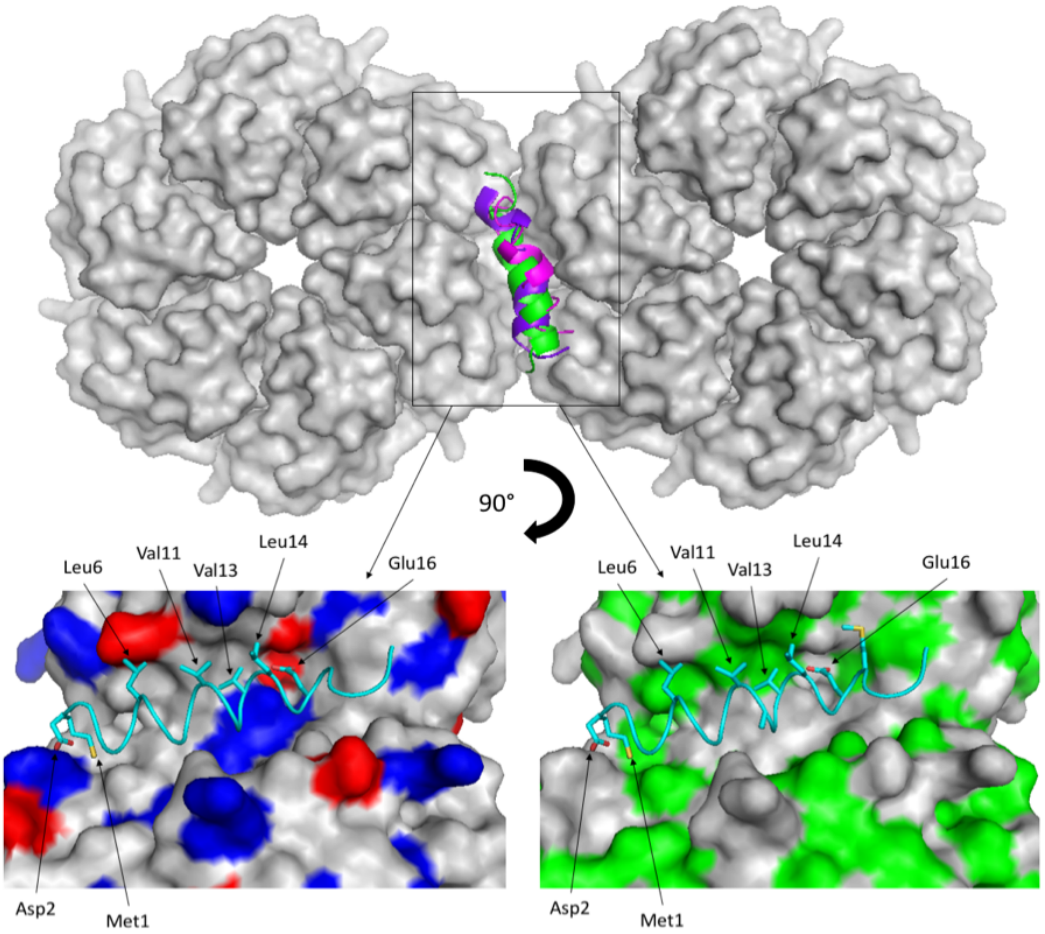

FIGURE 7.1: Model of the PduA- Signalling peptide complex, L20 is shown in green, P18 in pink and D18 in purple, all localise to the hexamer- hexamer interface. A zoomed view of the PduA- L20 complex rotated 90 degrees clockwise is also presented with the positive and negative residues of PduA coloured blue and red and the hydrophobics coloured green, in both images the conserved residues of L20 are shown as sticks, showing that the conserved, hydrophobic residues appear to bury themselves in a highly hydrophobic region of PduA.

The models produced (figure: 7.1) show all signalling peptides localised to the hexamer-hexamer interface. All peptides bind in an alpha helical conformation, consistent with the CD data reported earlier (figure: 6.5). The interaction is primarily hydrophobic, where Leucine and Valine residues on the peptide are buried in a poly-alanine sequence, which is present on the α- 2 helices of PduA. These hydrophobic residues within the peptide are conserved between PduL, P and D, and the poly-alanine sequence is largely conserved between hexameric shell proteins (figure: 7.2).

This would suggest that these residues play some important role in the microcompartment.

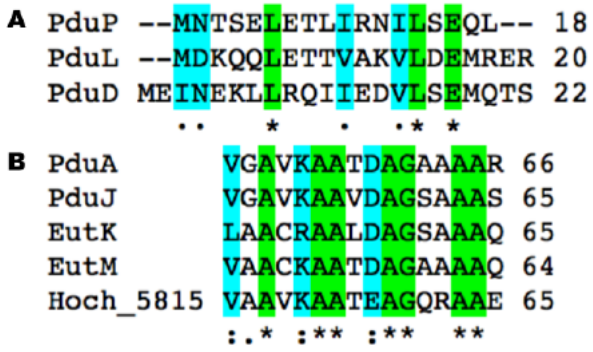

FIGURE 7.2: Alignment of known Pdu signalling peptides (A) and also the Alpha 2 helix of hexameric shell proteins (B). Fully conserved residues are highlighted green and partially conserved residues are highlighted blue. The regions stipulated to be involved in binding, in both the peptides and PduA are largely conserved.

The peptide binding at the interface is also able to explain the earlier observed four bands seen on Native PAGE, where PduA is not totally saturated (figure: 6.10). In chapter: 3, it was proposed that PduA assembles into trimers in solution, and when assessed using Native PAGE. If the peptide binds to the interface, this would present three potential binding sites and produce four potential species before total saturation, with 0, 1, 2 and 3 peptides bound (figure: 7.3).

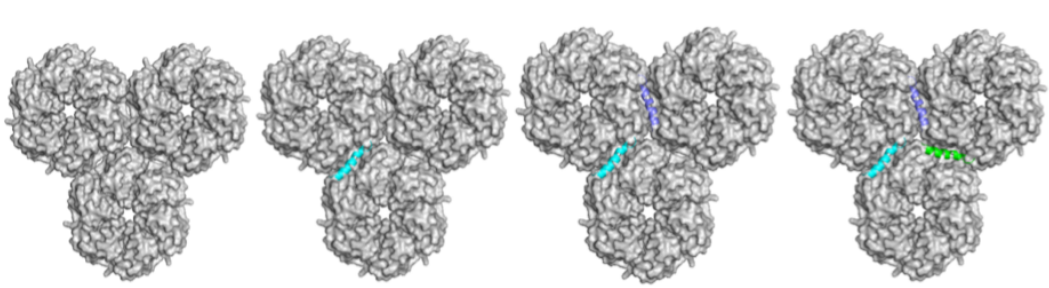

FIGURE 7.3: Proposed trimers of PduA hexamers with 0, 1, 2 and 3 L20 peptides bound at the interface between hexamers. This model accounts for the three potential species seen in figure: 7.3, as four different species would be present before total saturation of PduA.

7.1.1 Modelling the binding of PduL to curved assemblies of PduA.

In the previous chapter (section: 6.4), it was hypothesised that the binding of PduL to PduA sheets inhibited the formation of angles between tessellated hexamers. Consequently, this may even induce disassembly of curved assemblies (discussed in section: 6.4). Therefore, this hypothesis was tested using an in silico approach. Various PDB files were produced in pymol from a pair of PduA hexamers, featuring an increasing bend angle between the two hexamers (from 180- 140°). These PDB files were then used as an input model on the HPEPDOCK webserver [82], with the L20 peptide supplied as the ligand. As anticipated, where then angle between hexamers is reduced fewer peptides are targeted to the interface (figure: 7.4B). In instances where peptides are targeted to the interface, fewer are able to adopt the known alpha helical conformation.

The interface between hexamers is clearly the preferential binding site. However, where the angle between hexamers decreases (reducing the acccessability of this binding site), peptides are instead targeted to the C- terminus of PduA, on the opposing surface. This is the binding site that had previously been reported [51, 52] and is the binding site predicted here, only when the interfacial binding site is obscured.

Not only do the peptides show a clear preference for binding to the hexamer- hexamer interface, but in instances where this binding site is occluded, the average energy score of binding is lower (figure: 7.4C). The number of peptides bound at the interface, also appears to positively correlate with the average energy score of binding.

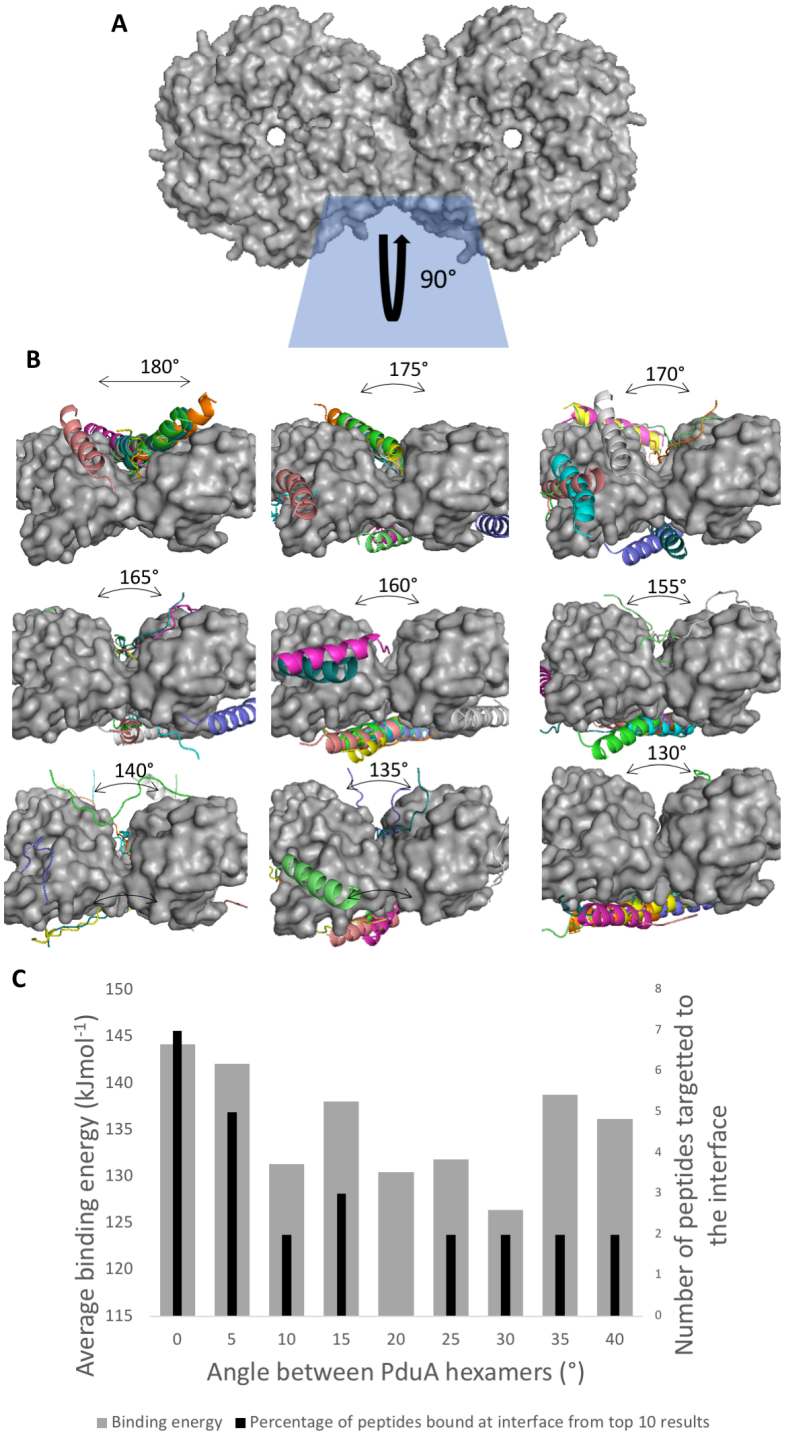

FIGURE 7.4: A top down view of a pair of PduA hexamers, clearly showing the groove present at the interface of two hexamers (A). Several zoomed in images are shown where A is rotated 90° degrees in the Z- plane (B) each of these images shows the top 10 scores of a HPEPDOCK docking simulation where the PduA hexamers have been rotated relative to each other in order to close the groove present, where smaller angles are present fewer L20 peptides target to the interface. A bar chart showing number of peptides targeted to the interface per top 10 models and the binding energy of these models (C) shows that where the angle between PduA molecules is reduced, fewer peptides are targeted to the interface. The average binding energy of the top 10 models is typically higher where bend angles are less and peptides are targeted to the interface.

7.1.2 Using the model to account for increased migration on Native PAGE.

As previously mentioned, the increased mobility of the PduA- PduL complex is unexpected. However, it can be accounted for if there is a sufficient change in the net negative charge of the complex. Using the model described (figure: 7.1), the crystal structure of PduL (section: 5.4) was superimposed onto the L20 peptide to give an estimate of the structure of the whole complex (figure: 7.5). The complex generated was satisfactory, with no visible clashes between PduA and PduL or individual PduL dimers. Additionally, the location of PduL with respect to the PduA trimer of hexamers demonstrates that three PduL dimers could feasibly bind to a trimer of PduA hexamers, in accordance with figure: 7.3.

Two parameters were derived from the model of the PduA- PduL complex:

1) The overall net charge of the complex. An electric field was calculated from the PDB file at the pH experienced within the context of Native PAGE (8.5). Calculating the charge from the 3- dimensional structure of a molecule is useful in the context of understanding the behaviour of a native molecule. Where charge is calculated from the polypeptide chain, every residue is given equal weighting and so any values calculated will relate to the unfolded protein rather than in its native state. The PDB2PQR web server [97] uses the Poisson- Boltzmann equation to determine the mean electric field of a PDB file, using only the solvent exposed residues. This gives a far better estimate of the charge of a protein in it's native state rather than deriving charge from sequence alone.

2) Hydrodynamic radius. Hydrodynamic radius was chosen as a measure rather than molecular weight as this better describes the mobility of a molecule when travelling through a medium such as a gel filtration column, or in this case an acrylamide gel matrix. PduA is an unusual protein, in that it due to its flatness it has a far larger radius than is typical for proteins of similar masses. Therefore hydrodynamic radius was used, in order to better compare differences in migration.

TABLE 7.1: Electrostatic parameters and hydrodynamic radii of the PduA- PduL complex. The negative nature of PduL, and the fact that positive residues are disproportionally buried following complex formation, results in an increased charge: mass ratio for the larger species. This would explain why the PduA- PduL complex travels further than PduA when assessed using native PAGE, despite the fact the complex is larger, the increased negativity promotes increased migration.

PduL type	pI (unfolded)	pI (folded)	Charge at pH 8.5	Hydrodynamic radius (A)	Charge/ Hydrodynamic radius
3 PduA	7.43	7.23	17.57	46.96	0.37
3 PduA + 1 X PduL	7.15	6.95	-27.41	54.04	0.51
3 PduA + 2 X PduL	6.99	6.81	-37.47	59.8	0.63
3 PduA + 3 X PduL	6.90	6.72	-47.42	65.51	0.72

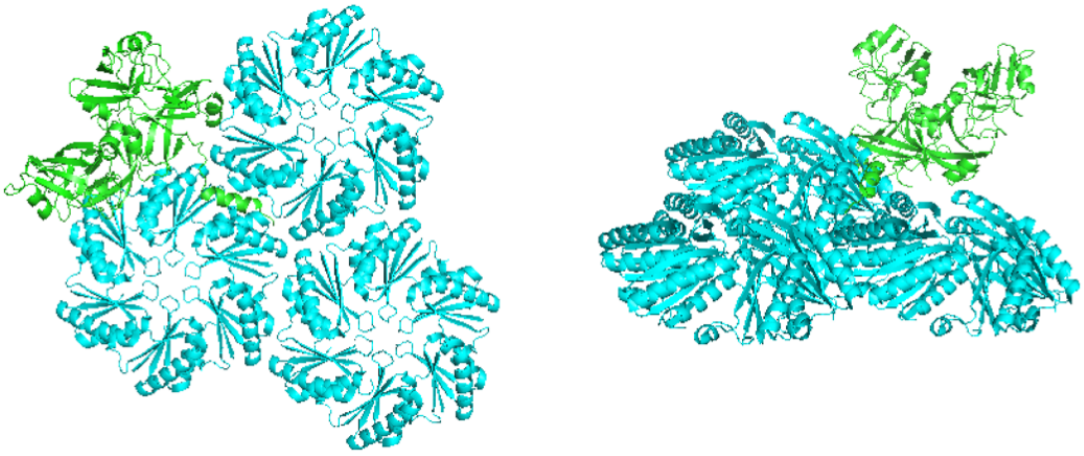

FIGURE 7.5: A Model of a trimer of PduAs (blue) binding a PduL dimer (green). This model was produced by aligning the known crystal structure of PduL (chapter: 5) with the model shown in figure: 7.1. Two angles are shown, demonstrating that there are no clashes between the two proteins. PduL sits on the convex face of PduA facing outwards, relative to the centre of the trimer of hexamers.

Table 7.1 shows various parameters calculated for the different species predicted for each band seen in the Native PAGE (figure: 6.10). As expected the pI for calculated for unfolded protein is higher than that for the native (folded protein). Additionally, whilst the size, and thereby hydrodynamic radius of each species increases, the net

negative charge also increases resulting in an increase in the charge: mass ratio. Consequently, this explains why species with larger molecular weights may migrate further if there is a sufficient change in overall charge.

7.2 Probing the PduA- PduL model using mutagenesis.

7.2.1 Trapping the complex through the introduction of a disulphide.

In an attempt to trap the predicted PduA- PduL complex, mutants were designed using the previously discussed model (figure: 7.1). The model was probed for suitably proximal residues that might form a disulphide bond if they were changed to cysteine. Asparagine 67 from PduA and aspartate 15 from PduL were selected as being suitably located (figure: 7.6) and the mutants D15C and N67C were produced using site directed mutagenesis.

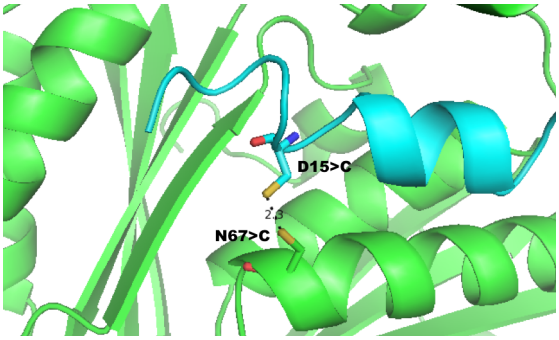

FIGURE 7.6: Model of the interaction proposed in figure 7.1 with D15 and N67 mutated to cysteine, showing they are suitably proximal for disulphide bond formation.

The purified mutants D15C and N67C were added in an equimolar ratio to one another and to WT PduA and L respectively. The mutant mixtures were then treated with reduced or oxidised glutathione and also with SDS and boiled. The resultant samples were run on Native PAGE. If a disulphide bond had formed between the mutants, a single band should be present in the sample which had been oxidised and denatured, due to the presence of a disulphide. Similarly, the lane which was reduced and denatured should contain two bands, as the disulphide had been reduced.

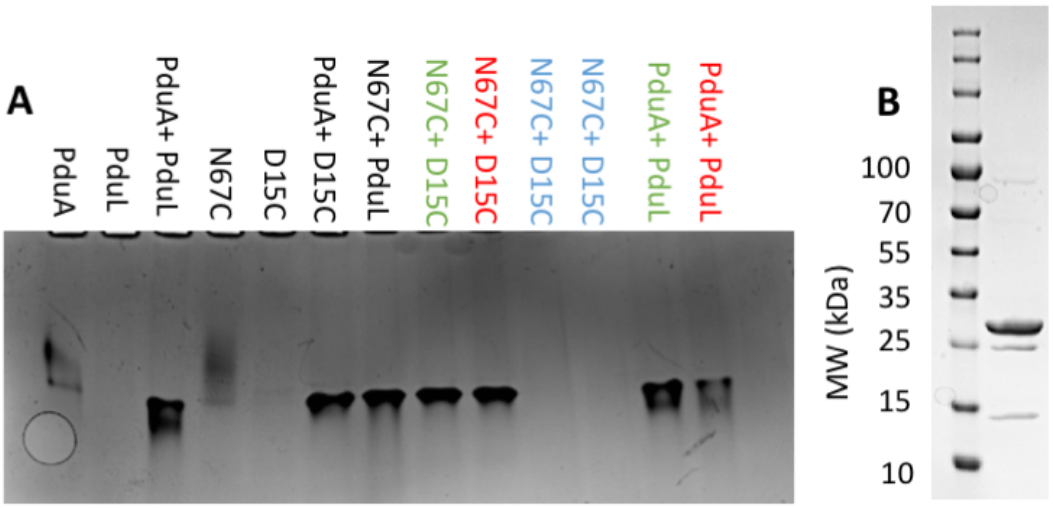

FIGURE 7.7: Native PAGE of PduA, PduL and mutants N67C and D15C (A). Native samples are labeled in black, oxidised samples in green, reduced samples in red and samples which have been both reduced and denatured in blue. Both cysteine mutants (N67C and D15C) were still able to interact with PduA, PduL and one another, demonstrating the cysteine did not inhibit complex formation. Where they were reduced the band persisted, suggesting the complex formed was not only mediated by a disulphide bond. Where the mutants were treated with SDS- and boiled, no band was apparent in either the oxidised or reduced lane. If a disulphide bond had formed a band should still be visible where the sample was oxidised but not reduced. However, as this was not the case no band was present. An SDS- PAGE analysis (with the omission of a reducing agent) (B) showed two bands corresponding to both D15C and N67C. If a disulphide had formed a band would be present at 36 kDa, corresponding to both proteins linked together. Interestingly, a band just below 26 kDa indicates that the PduL mutant, D15C has degraded. Such degradation does not occur in WT PduL, indicating the substitution of aspartate for cysteine has reduced the stability.

Both D15C and N67C migrated similarly to their WT counterparts. Most importantly, N67C appeared to tesselate (figure: 7.7), as evidenced by its similar migration to PduA. N67C and D15C were also found to form a complex with WT PduL and PduA respectively, indicating that the residues selected for mutation were not critical in complex formation. As expected the mutant complex ran as a single band and was not affected by the addition of either GSH or GSSH. When these oxidised and reduced samples were denatured by boiling in SDS no band was visible, suggesting the unfolded protein had run off the gel with the dye front. Samples containing N67C and D15C were then run on a standard SDS- PAGE, with the omission of BME

in the loading buffer so as to permit the formation of disulphides. If a disulphide had formed, a mass at 36 kDa should be present. However, there was no band at 36 kDa and masses at 12 and 24 kDa were present, corresponding to both free PduA and PduL.

Both sequences confirmed the mutagenesis was successful. Additionally the presence of a free cysteine on N67C was confirmed by the addition of DTNB (which binds to free cysteines), which produced a bright yellow colour change when added to N67C but not WT PduA (figure: 7.8). Despite the mutagenesis being successful, there was a lack of cross- linking between mutants. Assuming the model produced is reasonably correct, both asparagine 67 and aspartate 15 should have been suitably proximal to promote disulphide formation. However, the angle present between the two cysteines may not have been suitable for the formation of a disulphide. Disulphide formation has a clear preference towards dihedral angles of 90° relative to the two sulphurs. Figure: 7.6 clearly shows that the cysteines present at positions 15 and 67 are not located 90° relative to one another. Additionally, despite having sufficient evidence that the L20 peptide binds at the hexamer- hexamer interface, the precise localisation of the peptide has only been estimated using simulations. Whilst all software used has targeted L20 to the hexamer- hexamer interface, there are slight variations as to the exact position of the peptide. It is therefore likely that whilst the peptide does bind to the hexamer- hexamer interface, the position of asparagine 67 and aspartate 15 as determined from the model (figure: 7.6) are not precise. Consequently, the position of these residues are not conducive to the formation of a disulphide.

FIGURE 7.8: PduA (left) and PduA N67C (right), DTNB was added to both samples, no colour change is observed following addition to PduA however, the free cysteine introduced in N67C produces a colour change to a vivid yellow.

7.2.2 Disrupting the interaction through introduction of an Arginine.

The previous attempt to demonstrate PduL binds to PduA at the interface between two hexamers, by introducing a disulphide bond was unsuccessful. Therefore, a less specific approach was required. Here, an arginine was introduced at position 63 in place of alanine on the α2 helix of PduA (A63R). The presence of a large bulky residue such as arginine in the cleft between two PduA hexamers should restrict access to the cleft, whilst still allowing tessellation of the PduA hexamers. The previous technique, using a predicted complex to insert a disulphide bond required the precise positioning of both thiol groups in order to trap the complex. However, whilst this method is less precise, it should be able to ascertain whether the extension peptide of PduL binds at the cleft, as hypothesised.

Initial attempts to assess the interaction of PduA (A63R) and PduL using Native PAGE were unsuccessful. Surprisingly, despite the concentration of A63R being confirmed by SDS- PAGE and A_{280} absorbance, no band was detectable using Native PAGE. Subsequent Native PAGE analysis were performed using higher concentrations of A63R. However, there was still no band indicating the presence of the protein in the gel. Analysis of the isoelectric points of both PduA and the A63R variants provided the answer to this puzzle. The substitution of alanine for arginine had increased the net positive charge of the protein. A comparison between the predicted pI of PduA and A63R revealed a change of 7.23 to 9.06 when arginine is introduced.

Native PAGE is typically run at a pH of 8.5, for wild type PduA with a pI of 7.23, this results in a net negative charge on the protein (as the pI is less than the pH of the running buffer). However, for A63R the pI of 9.06 would result in a positive charge at pH 8.5 and result in the protein being actively repelled from the gel by the positive electric field. To counter this a Native gel was poured at pH 9.5 and the pH of the running and loading buffers was adjusted to pH 9.5. Here, A63R was able to migrate into the gel and could be visualised.

Native PAGE analysis at pH 9.5 was performed on PduA, PduA$_{GST}$ and A63R with the addition of PduL (figure: 7.9). This revealed similar results for both PduA and PduA$_{GST}$ to those seen at pH 8.5, indicating that complex formation occurs in the case of PduA and not with PduA$_{GST}$. Although, PduL does run further due to the increase in pH producing an increase in net negative charge. The migration of A63R however, was unaffected and the band corresponding to PduL persisted and was similar in intensity. This suggested that the arginine succeeded in disrupting the binding of PduL, proving that the signalling peptide of PduL does indeed bind at the interface of two PduA hexamers.

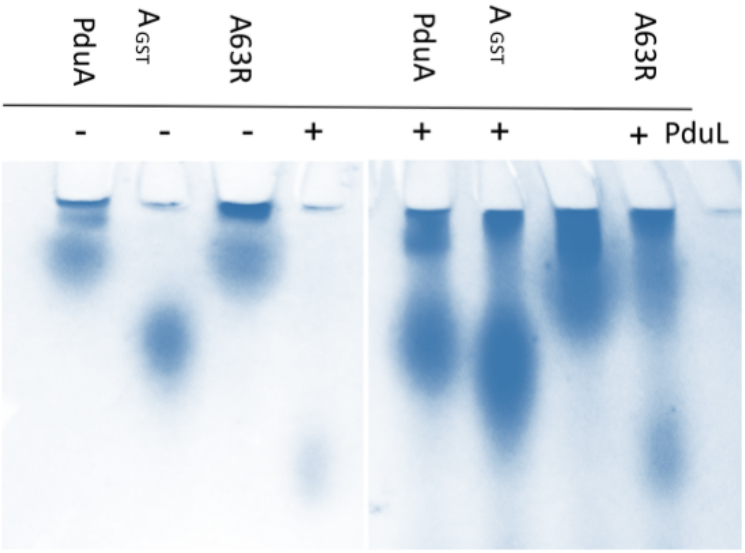

FIGURE 7.9: Native PAGE of PduA, PduA$_{GST}$ and A63R in the presence and absence of PduL. Bands for PduL and A63R are unaltered in position or intensity following their combining, showing that PduL is unable to interact with the A63R mutant, and thereby demonstrating that the L20 peptide does bind at the interface between two PduA hexamers, and that when this site is occluded by the introduction of arginine, no binding is detectable.

7.2.3 Using the partially tessellating PduA mutant to demonstrate the significance of hexamer tessellation in PduL binding.

The mutant described in chapter: 3, section: 3.5, $K26D_4$ had shown a decreased propensity to tesselate. Partially tessellated species (monomers and trimers of hexamers), were also detectable using Native PAGE. This feature represented a unique opportunity to demonstrate that tessellation is necessary for the interaction with PduL. Only the migration of the top band which corresponds to the tessellating protein should be affected by the addition of PduL. If PduL does bind at the interface between hexamers then the lower band, relating to a single hexamer should remain unaffected.

Native PAGE was performed on an aged sample of the $K26D_4$ protein (figure: 7.10). This sample had significant populations of both multimers and monomers of PduA hexamers (Ascertained from the presence of two bands in a significant amount). Here, following the addition of PduL only the migration of the top band was effected whilst the bottom band remained the same in both position and intensity. This experiment was therefore able to demonstrate that out of two populations of sequentially identical PduA mutants, only the multimeric form was able to interact with PduL. Thereby demonstrating the critical importance of the tessellation of PduA in the binding of PduL.

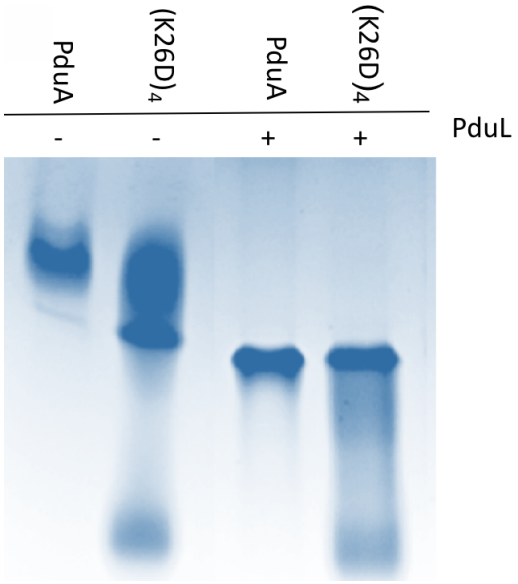

FIGURE 7.10: Native PAGE of PduA (K26D)4(WT)2 with PduL. This PduA mutant tessellates slowly over time, and only the migration of the tessellating protein (top band) is affected by the addition of PduL. The position and intensity of the bottom band (which corresponds to a single hexamer) remained unchanged following the addition of PduL.

7.3 Conclusions and further work.

The combination of in silico docking and mutagenic experiments described in this chapter indicate that the N- terminus of PduL binds to the interface between two PduA hexamers. This model is able to reconcile all experimental evidence within this thesis and is also in accordance with current thought regarding the orientation of BMC shell proteins.

The synthesis of artificial signalling peptides, with varying affinities would be of particular benefit to industrial applications, as it would allow some control of the amount of specific enzymes to be packaged into the microcompartment. However, in order for this to be feasible the manner in which native signalling peptides bind must be known. Here, a model has been presented showing the peptide binding at the hexamer- hexamer interface. Whilst there is sufficient evidence to suggest this is the case, the precision of the model is not certain (as evidenced by the attempted disulphide formation, section: 7.2.1). Further mutagenesis, removing the

specific residues speculated to be involved in the interactions would be beneficial. This would allow for the elucidation of key interactions, rather than restricting access to the binding site altogether. Information on the binding constants would also be of great benefit in order to ascertain whether certain mutations effect the affinity of peptide to shell protein, even if they do not totally abolish it. Ideally, crystallographic data would give a precise model of where the peptide binds and the specific residues involved. Whilst this has been attempted multiple times there are various factors which may have influenced the lack of success, as detailed below.

Crystallography was attempted although the complex was never successfully co-crystallised, both PduA and PduL were crystallised in isolation despite both proteins being present in the crystallisation mix. PduA crystals were also produced and soaked in the a solution containing the signalling peptides and well solution, although when this was attempted the crystals were seen to crack and dissolve. Anecdotally, this is further evidence for the peptide binding at the hexamer- hexamer interface, as the peptide is disturbing the crystal contacts resulting in its degradation. The orientation of the peptide in the binding site may also limit the scope of crystallography. The interface where the peptide binds is symmetrical when rotated through 180 °, and there are six interfaces present within the unit cell. This results in there being 64 symmetrically unique orientations of the peptide in a single unit cell, which would cancel each other out and not be detectable in the final electron density map.

Chapter 8

Conclusions.

8.1 Conclusions and discussions.

Throughout the past 60 years since the initial observations of polyhedral bodies in the cytoplasm of photosynthetic bacteria, much has been done to elucidate the roles and mechanisms of BMCs and the functions they perform. Despite this there have still been several areas of research where insufficient understanding has hindered progress. The manner in which shell proteins assemble into complete microcompartments is one such area, although there has been several breakthroughs regarding this in recent years [18, 85], which have begun to elucidate the assembly processes undertaken during the formation of microcompartments.

8.1.1 Small oligomers of shell protein hexamers preclude the formation of higher order structures.

The work completed here in chapter: 3 has illuminated the assembly of small oligomers of PduA, which likely preclude the formation of the larger facets of the flat faces of the BMC. Additionally, PduK was found to exist in equilibrium between small oligomers and single hexamers (section: 6.5). What is most surprising about these intermediates is the speed at which they assemble, where freshly purified samples of PduA are show to be entirely multimeric. These small oligomers are relatively stable once formed, although the same sample will increase in mass when tested repeatedly over a period of days. This indicates that further oligomerisation occurs

more slowly than that of the initially formed species. How these small oligomers relate to the structure and formation of whole BMCs is unclear, although they may form the basis of the scaffolds to which other shell proteins adhere. This explanation is supported by the recent structure of a complete microcompartment [85], where the only shell protein to interact with itself is a protein, highly homologous to PduA. All other proteins in this system interact exclusively with this PduA like protein. Evidence showing PduA to be one of the most abundant shell protein in the Pdu microcompartment [18] also supports this theory.

8.1.2 The signalling peptide present on the N- terminus of PduL binds to the interface between PduA hexamers.

The discovery of small, soluble PduA oliogomers raised further questions, regarding where the extension peptides present on the N- termini of encapsulated enzymes within the microcompartment localise to. It had previously been thought that these peptides targeted enzymes to the C- terminus of select shell proteins [48, 52, 98] resulting in their encapsulation. However, data showing that the C- termini were present on the surface of shell proteins presented on the exterior of microcompartments [30, 85] brought this idea into dispute. The results explored in chapter: 3 presented an opportunity to investigate whether these extension peptides bound to the cleft present between adjacent shell proteins, as had been previously speculated [94]. There were variants of PduA available which exhibited different propensities towards tessellation discussed in chapter: 3. The interaction between these mutants and the encapsulated enzyme, PduL, demonstrated that where non- tessellating variants of PduA were used, no interaction could be detected with PduL (chapter: 6, section: 6.3). Following this discovery, models were produced localising the PduL signalling peptide to the interface between adjacent PduA hexamers (chapter: 7). Where a pair, or trio of hexamers were provided as the input for the modelling software, the binding site was predicted to be at the interface between hexamers an overwhelming majority of the time, regardless of the software used. The validity of these models was explored using site directed mutagenesis. Whilst

the model could be confirmed as being precisely correct, the binding site at the interface of two hexamers was confirmed as the true binding site. This was achieved by the introduction of an arginine at the PduA interface (chapter: 7, section: 7.2.2). This mutant obstructed access to the binding site, yet was still capable of tessellation. Consequently, this demonstrated the significance of this cleft in binding rather than some other element introduced when shell proteins oligomerise.

The discovery that L20 binds at the interface between two hexamers is significant. With knowledge of how the encapsulation peptides bind, and the location of where they bind to the microcompartment, synthetic biology applications of BMCs can be more precisely engineered. As in nature, various components within the BMC catalyse reactions at different rates. Therefore, in order to maintain some level of homeostasis in regards to the small molecules present within a microcompartment, the enzymes within must be encapsulated to varying degrees. This can be observed where the constituents of a microcomparmtent are analysed, not all enzymes are present in equal proportions [27, 99]. If the encapsulation peptides are entirely responsible for the encapsulation of the enzymes, then the slight deviations in the sequence of these peptides must produce varying affinities. Consequently, this allows for differential levels of encapsulation of different enzymes. This natural occurrence could be manipulated in order to achieve varying levels of encapsulation of non- native enzymes, something which is desirable where multiple enzymes in a pathway have varying catalytic activities. There are several industrially relevant biochemical pathways which may benefit from encapsulation within a BMC and these pathways would require different levels of encapsulation of their components. Knowledge of how the signalling peptides interact with the compartments, and the individual strengths of those interactions would be necessary to facilitate this differential packing. Artificial signalling peptides with varying affinities could also be synthesised, allowing for further optimisation.

8.1.3 Substrate transfer between encapsulated enzymes within the Pdu BMC.

In chapter: 6, section: 6.6.2 for the first time, direct evidence of enhanced substrate transfer between constituent enzymes of a BMC was observed. In this instance, flux of CoA between PduP and PduL was increased in the presence of PduA. Other studies have demonstrated the importance of the BMC shell for maintaining internal substrate concentrations and thereby maintaining the efficiency of the enzymes within. The experiments described here showed that a significant increase in substrate transfer is observable, even when the enzymes are not present together in an enclosed compartment. From this it can be inferred that by localising enzymes to the same small PduA oligomer a change in substrate transfer is elicited merely by having the two enzymes become more proximal, thereby limiting the diffusion distance of CoA. This effect is likely also present within the microcompartment and perhaps the signalling peptides play a role in co- localising enzymes which feature coupled reactions.

8.1.4 The role of the C- terminus in BMC shell proteins.

The function of the C- terminal extension present on a number of shell proteins has been a persistent mystery in the study of BMCs. Here, an investigation in to the C- terminus of PduK (chapter: 3, section: 3.6) found that the C- terminal domain influences both the proteins ability to form higher order structures and how it interacts with PduL. A mutant lacking this domain was found to not form the large, flat sheets formed by WT PduK, and instead formed objects more homologous with the tubes formed by PduA. This truncated mutant also had a substantially lower affinity for PduL than its full length counterpart, as assessed by fluorescence quenching (section: 6.5.1). This result would seem to be in accordance with the theory that the C- terminus of PduA binds to PduP [48] and therefore, perhaps the C- terminus of PduK serves a similar role. However, as has been repeatedly stated, the C- terminus of these hexameric proteins is present on the exterior of BMCs. Therefore, this theory conflicts with current evidence regarding the orientation of shell proteins. The

observation that truncated PduK forms tube like objects does however support the notion that the signalling peptides bind at the hexamer- hexamer interface. Where shell proteins form tubes and compartments, it is intuitive that some curvature must be present at the interface between hexamers. Therefore, where the C- terminus of PduK facilitates the formation of flat sheets rather than curved objects, the proposed binding site at the hexamer- hexamer interface is more readily accessible by PduL. Consequently, this is what produces the tighter binding where the C- terminus is present, rather than a direct interaction between PduL and the C- terminus itself. If the C- terminus of PduA serves a similar function then this reconciles the previously reported decrease in affinity seen between certain PduA C- terminal mutants and PduP. Rather than the C- terminus playing a direct role in binding to the enzyme, it may induce a conformational change by limiting curvature and increasing access to the true binding site. This binding site is present on the opposite surface of the hexamer with respect to the C- terminus, and the planar conformation of tessellating shell proteins is more conducive to binding.

8.1.5 Concluding remarks.

The study of BMCs has evolved rapidly in recent years, and many of the fundamental questions regarding assembly and cargo encapsulation have been illuminated. Further work to establish exactly how these processes work, and how they deviate between different microcompartments still requires attention. In particular, a mode of binding where signalling peptides bind at the interface between two hexamers has been presented here. However, this is not necessarily the single mechanism by which enzymes are encapsulated. In addition to the groove present at the interface of tessellating hexamers, a significant cleft is present at the interface of adjacent monomers within the PduA hexamer. This groove may be an alternative binding site, although the fact that PduL does not appear to interact with non- tessellating PduA mutants conflicts with this. The nature of the interface between different shell proteins is also relatively unknown, as crystal structures are typically of a homogenous nature. The size, curvature and polarity of these interfaces must vary depending on which shell proteins are interacting. Consequently, different interfaces may preferentially

interact with different signalling peptides, producing a differential distribution of enzymes throughout the microcompartment. Further knowledge of these preferences would be invaluable in building a more complete picture of how microcompartments assemble and a thorough model of the Pdu BMC. In conclusion, a general understanding is emerging that the binding of enzymes to one another precludes the formation of the outer shell of the microcompartment [53]. These shells are composed, predominantly of a single type of shell protein which functions as a scaffold to tether the other shell proteins and enzymes to one another. The work described in this thesis underpins and supports these principles, while providing further evidence for how these shell protein scaffolds form and how the encapsulated enzymes bind at the interface between tessellating shell proteins, which facilitates the formation of the flat facets of the microcompartment.

Bibliography

[1] Drews, G. & Niklowitz, W. Cytology of cyanophycea. ii. centroplasm and granular inclusions of phormidium uncinatum. *Archiv Fur Mikrobiologie* **24**, 147 (1956).

[2] Lang, N. J. The fine structure of blue-green algae. *Annual Reviews in Microbiology* **22**, 15–46 (1968).

[3] Shively, J., Ball, F., Brown, D. & Saunders, R. Functional organelles in prokaryotes: polyhedral inclusions (carboxysomes) of thiobacillus neapolitanus. *Science* 584–586 (1973).

[4] Shively, J., Ball, F. L. & Kline, B. W. Electron microscopy of the carboxysomes (polyhedral bodies) of thiobacillus neapolitanus. *Journal of Bacteriology* **116**, 1405–1411 (1973).

[5] Price, G. D., Coleman, J. R. & Badger, M. R. Association of carbonic anhydrase activity with carboxysomes isolated from the cyanobacterium synechococcus pcc7942. *Plant Physiology* **100**, 784–793 (1992).

[6] Price, G. D. & Badger, M. R. Evidence for the role of carboxysomes in the cyanobacterial co2-concentrating mechanism. *Canadian Journal of Botany* **69**, 963–973 (1991).

[7] Reinhold, L., Kosloff, R. & Kaplan, A. A model for inorganic carbon fluxes and photosynthesis in cyanobacterial carboxysomes. *Canadian journal of botany* **69**, 984–988 (1991).

[8] Kerfeld, C. A. *et al.* Protein structures forming the shell of primitive bacterial organelles. *Science* **309**, 936–938 (2005).

[9] Tanaka, S. *et al.* Atomic-level models of the bacterial carboxysome shell. *science* **319**, 1083–1086 (2008).

[10] Cai, F. *et al.* The pentameric vertex proteins are necessary for the icosahedral carboxysome shell to function as a co 2 leakage barrier. *PloS one* **4**, e7521 (2009).

[11] Jeter, R. M. Cobalamin-dependent 1, 2-propanediol utilization by salmonella typhimurium. *Microbiology* **136**, 887–896 (1990).

[12] Bobik, T. A., Ailion, M. & Roth, J. R. A single regulatory gene integrates control of vitamin b12 synthesis and propanediol degradation. *Journal of Bacteriology* **174**, 2253–2266 (1992).

[13] Chen, P., Andersson, D. I. & Roth, J. R. The control region of the pdu/cob regulon in salmonella typhimurium. *Journal of bacteriology* **176**, 5474–5482 (1994).

[14] Bobik, T. A., Havemann, G. D., Busch, R. J., Williams, D. S. & Aldrich, H. C. The propanediol utilization (pdu) operon ofsalmonella enterica serovar typhimurium lt2 includes genes necessary for formation of polyhedral organelles involved in coenzyme b12-dependent 1, 2-propanediol degradation. *Journal of bacteriology* **181**, 5967–5975 (1999).

[15] Havemann, G. D., Sampson, E. M. & Bobik, T. A. Pdua is a shell protein of polyhedral organelles involved in coenzyme b12-dependent degradation of 1, 2-propanediol in salmonella enterica serovar typhimurium lt2. *Journal of Bacteriology* **184**, 1253–1261 (2002).

[16] Chowdhury, C. *et al.* Selective molecular transport through the protein shell of a bacterial microcompartment organelle. *Proceedings of the National Academy of Sciences* **112**, 2990–2995 (2015). URL http://www.pnas.org/content/112/10/2990.abstract.

[17] Crowley, C. S., Sawaya, M. R., Bobik, T. A. & Yeates, T. O. Structure of the PduU Shell Protein from the Pdu Microcompartment of Salmonella. *Structure* **16**, 1324–1332 (2008).

[18] Parsons, J. B. *et al.* Synthesis of Empty Bacterial Microcompartments, Directed Organelle Protein Incorporation, and Evidence of Filament-Associated

Organelle Movement. *Molecular Cell* **38**, 305–315 (2010). URL http://dx.doi.org/10.1016/j.molcel.2010.04.008.

[19] Cheng, S., Sinha, S., Fan, C., Liu, Y. & Bobik, T. A. Genetic Analysis of the Protein Shell of the Microcompartments Involved in Coenzyme B12-Dependent 1,2-Propanediol Degradation by Salmonella. *Journal of bacteriology* **193**, 1385–92 (2011). URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3067621{&}tool=pmcentrez{&}rendertype=abstract.

[20] Pang, A., Warren, M. J. & Pickersgill, R. W. Structure of PduT, a trimeric bacterial microcompartment protein with a 4Fe-4S cluster-binding site. *Acta Crystallographica Section D: Biological Crystallography* **67**, 91–96 (2011).

[21] Pang, A., Liang, M., Prentice, M. B. & Pickersgill, R. W. Substrate channels revealed in the trimeric Lactobacillus reuteri bacterial microcompartment shell protein PduB. *Acta Crystallographica Section D: Biological Crystallography* **68**, 1642–1652 (2012).

[22] Parsons, J. B. *et al.* Biochemical and structural insights into bacterial organelle form and biogenesis. *Journal of biological chemistry* **283**, 14366–14375 (2008).

[23] Lehman, B. P., Chowdhury, C. & Bobik, T. A. The n terminus of the pdub protein binds the protein shell of the pdu microcompartment to its enzymatic core. *Journal of bacteriology* **199** (2017).

[24] Crowley, C. S. *et al.* Structural insight into the mechanisms of transport across the Salmonella enterica Pdu microcompartment shell. *Journal of Biological Chemistry* **285**, 37838–37846 (2010).

[25] Pang, A., Frank, S., Brown, I., Warren, M. J. & Pickersgill, R. W. Structural insights into higher-order assembly and function of the bacterial microcompartment protein PduA. *The Journal of biological chemistry* **289**, 22377–22384 (2014). URL http://www.ncbi.nlm.nih.gov/pubmed/24873823.

[26] Chowdhury, C., Chun, S., Sawaya, M. R., Yeates, T. O. & Bobik, T. A. The function of the pduj microcompartment shell protein is determined by the genomic position of its encoding gene. *Molecular microbiology* **101**, 770–783 (2016).

[27] Yang, M. *et al.* Decoding the stoichiometric composition and organisation of bacterial metabolosomes. *Nature communications* **11**, 1–11 (2020).

[28] Tanaka, S., Sawaya, M. R. & Yeates, T. O. Structure and mechanisms of a protein-based organelle in escherichia coli. *Science* **327**, 81–84 (2010).

[29] Lawrence, A. D. *et al.* Solution structure of a bacterial microcompartment targeting peptide and its application in the construction of an ethanol bioreactor. *ACS synthetic biology* **3**, 454–465 (2014).

[30] Uddin, I., Frank, S., Warren, M. J. & Pickersgill, R. W. A Generic Self-Assembly Process in Microcompartments and Synthetic Protein Nanotubes. *Small* **14**, 1–7 (2018).

[31] Sutter, M., Greber, B., Aussignargues, C. & Kerfeld, C. A. Assembly principles and structure of a 6.5-mda bacterial microcompartment shell. *Science* **356**, 1293–1297 (2017).

[32] Fernández-Briera, A. & Garrido-Pertierra, A. A degradation pathway of propionate in salmonella typhimurium lt-2. *Biochimie* **70**, 757–768 (1988).

[33] Penrod, J. T. & Roth, J. R. Conserving a volatile metabolite: a role for carboxysome-like organelles in salmonella enterica. *Journal of Bacteriology* **188**, 2865–2874 (2006).

[34] Bobik, T. A., Xu, Y., Jeter, R. M., Otto, K. E. & Roth, J. R. Propanediol utilization genes (pdu) of salmonella typhimurium: three genes for the propanediol dehydratase. *Journal of Bacteriology* **179**, 6633–6639 (1997).

[35] Sauvageot, N., Muller, C., Hartke, A., Auffray, Y. & Laplace, J.-M. Characterisation of the diol dehydratase pdu operon of lactobacillus collinoides. *FEMS microbiology letters* **209**, 69–74 (2002).

[36] Shibata, N. *et al.* A new mode of b12 binding and the direct participation of a potassium ion in enzyme catalysis: X-ray structure of diol dehydratase. *Structure* **7**, 997–1008 (1999).

[37] Parsons, J. B. *et al.* Characterisation of pdus, the pdu metabolosome corrin reductase, and evidence of substructural organisation within the bacterial microcompartment. *PLoS ONE* **5** (2010).

[38] Johnson, C. L. *et al.* Functional genomic, biochemical, and genetic characterization of the salmonella pduo gene, an atp: cob (i) alamin adenosyltransferase gene. *Journal of bacteriology* **183**, 1577–1584 (2001).

[39] Mori, K. & Toraya, T. Mechanism of reactivation of coenzyme b12-dependent diol dehydratase by a molecular chaperone-like reactivating factor. *Biochemistry* **38**, 13170–13178 (1999).

[40] Bobik, T. A. Polyhedral organelles compartmenting bacterial metabolic processes. *Applied Microbiology and Biotechnology* **70**, 517–525 (2006).

[41] Cheng, S., Fan, C., Sinha, S. & Bobik, T. A. The PduQ Enzyme Is an Alcohol Dehydrogenase Used to Recycle NAD+ Internally within the Pdu Microcompartment of Salmonella enterica. *PLoS ONE* **7**, 1–11 (2012).

[42] Leal, N. A., Havemann, G. D. & Bobik, T. A. PduP is a coenzyme-a-acylating propionaldehyde dehydrogenase associated with the polyhedral bodies involved in B12-dependent 1,2-propanediol degradation by Salmonella enterica serovar Typhimurium LT2. *Archives of Microbiology* **180**, 353–361 (2003).

[43] Liu, Y. *et al.* Pdul is an evolutionarily distinct phosphotransacylase involved in b12-dependent 1, 2-propanediol degradation by salmonella enterica serovar typhimurium lt2. *Journal of bacteriology* **189**, 1589–1596 (2007).

[44] Liu, Y., Jorda, J., Yeates, T. & Bobik, T. The PduL phosphotransacylase is used to recycle coenzyme A within the Pdu microcompartment. *J Bacteriol* **197**, 2392–2399 (2015). URL http://jb.asm.org/lookup/doi/10.1128/JB.00056-15.

[45] Palacios, S., Starai, V. J. & Escalante-Semerena, J. C. Propionyl coenzyme a is a common intermediate in the 1, 2-propanediol and propionate catabolic pathways needed for expression of the prpbcde operon during growth of salmonella enterica on 1, 2-propanediol. *Journal of bacteriology* **185**, 2802–2810 (2003).

[46] Klein, M. G. *et al.* Identification and structural analysis of a novel carboxysome shell protein with implications for metabolite transport. *Journal of molecular biology* **392**, 319–333 (2009).

[47] Kerfeld, C. A. & Erbilgin, O. Bacterial microcompartments and the modular construction of microbial metabolism. *Trends in microbiology* **23**, 22–34 (2015).

[48] Fan, C. *et al.* Short N-terminal sequences package proteins into bacterial microcompartments. *Proc Natl Acad Sci U S A* **107**, 7509–7514 (2010). URL http://www.ncbi.nlm.nih.gov/pubmed/20308536.

[49] Fan, C. & Bobik, T. A. The n-terminal region of the medium subunit (pdud) packages adenosylcobalamin-dependent diol dehydratase (pducde) into the pdu microcompartment. *Journal of Bacteriology* **193**, 5623–5628 (2011).

[50] Liu, Y. *Role of the PduL enzyme in the function of primitive bacterial organelles found in Salmonella*. Ph.D. thesis (2014).

[51] Fan, C., Cheng, S., Sinha, S. & Bobik, T. A. Interactions between the termini of lumen enzymes and shell proteins mediate enzyme encapsulation into bacterial microcompartments. *Proceedings of the National Academy of Sciences* **109**, 14995–15000 (2012).

[52] Jorda, J., Liu, Y., Bobik, T. A. & Yeates, T. O. Exploring Bacterial Organelle Interactomes: A Model of the Protein-Protein Interaction Network in the Pdu Microcompartment. *PLoS Computational Biology* **11**, 1–23 (2015).

[53] Cameron, J. C., Wilson, S. C., Bernstein, S. L. & Kerfeld, C. A. Biogenesis of a bacterial organelle: the carboxysome assembly pathway. *Cell* **155**, 1131–1140 (2013).

[54] Huber, I. *et al.* Construction of recombinant pdu metabolosome shells for small molecule production in corynebacterium glutamicum. *ACS Synthetic Biology* **6**, 2145–2156 (2017).

[55] Wagner, H. J., Capitain, C. C., Richter, K., Nessling, M. & Mampel, J. Engineering bacterial microcompartments with heterologous enzyme cargos. *Engineering in Life Sciences* **17**, 36–46 (2017).

[56] Dadswell, K. *et al.* Bacterial microcompartment-mediated ethanolamine metabolism in escherichia coli urinary tract infection. *Infection and immunity* **87**, e00211–19 (2019).

[57] Chowdhury, C., Sinha, S., Chun, S., Yeates, T. O. & Bobik, T. A. Diverse bacterial microcompartment organelles. *Microbiology and Molecular Biology Reviews* **78**, 438–468 (2014).

[58] Kirst, H. & Kerfeld, C. A. Bacterial microcompartments: catalysis-enhancing metabolic modules for next generation metabolic and biomedical engineering. *BMC biology* **17**, 79 (2019).

[59] Prentice, M., Warren, M. & Liang, M. Accumulation of metabolic products in bacterial microcompartments (2015). US Patent 9,187,766.

[60] Kerfeld, C. A., Lassila, J. K. & Kinney, J. N. Constructs and systems and methods for producing microcompartments (2015). US Patent App. 14/214,172.

[61] Veronese, F. M. & Mero, A. The impact of pegylation on biological therapies. *BioDrugs* **22**, 315–329 (2008).

[62] Garg, V. K. *et al.* Mfppi–multi fasta protparam interface. *Bioinformation* **12**, 74 (2016).

[63] Mahmood, T. & Yang, P.-C. Western blot: technique, theory, and trouble shooting. *North American journal of medical sciences* **4**, 429 (2012).

[64] Dreisewerd, K., Schürenberg, M., Karas, M. & Hillenkamp, F. Influence of the laser intensity and spot size on the desorption of molecules and ions in matrix-assisted laser desorption/ionization with a uniform beam profile. *International journal of mass spectrometry and ion processes* **141**, 127–148 (1995).

[65] Steinberg, T. H., Haugland, R. P. & Singer, V. L. Applications of sypro orange and sypro red protein gel stains. *Analytical biochemistry* **239**, 238–245 (1996).

[66] Giacovazzo, C. *et al. Fundamentals of crystallography*, vol. 7 (Oxford University Press Oxford, 2002).

[67] Winter, G. xia2: an expert system for macromolecular crystallography data reduction. *Journal of applied crystallography* **43**, 186–190 (2010).

[68] Leslie, A. G. & Powell, H. R. Processing diffraction data with mosflm. In *Evolving methods for macromolecular crystallography*, 41–51 (Springer, 2007).

[69] Adams, P. D. *et al.* Phenix: a comprehensive python-based system for macromolecular structure solution. *Acta Crystallographica Section D: Biological Crystallography* **66**, 213–221 (2010).

[70] Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D: Biological Crystallography* **60**, 2126–2132 (2004).

[71] Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *Journal of molecular biology* **372**, 774–797 (2007).

[72] Davis, I. W. *et al.* Molprobity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic acids research* **35**, W375–W383 (2007).

[73] DeLano, W. L. Pymol (2002).

[74] Pennycook, S. J. & Nellist, P. D. *Scanning transmission electron microscopy: imaging and analysis* (Springer Science & Business Media, 2011).

[75] Greenfield, N. J. Circular dichroism (cd) analyses of protein-protein interactions. In *Protein-Protein Interactions*, 239–265 (Springer, 2015).

[76] Louis-Jeune, C., Andrade-Navarro, M. A. & Perez-Iratxeta, C. Prediction of protein secondary structure from circular dichroism using theoretically derived spectra. *Proteins: Structure, Function, and Bioinformatics* **80**, 374–381 (2012).

[77] Crouch, S. R. & Malmstadt, H. Mechanistic investigation of molybdenum blue method for determination of phosphate. *Analytical Chemistry* **39**, 1084–1089 (1967).

[78] Sabet-Azad, R., Linares-Pastén, J. A., Torkelson, L., Sardari, R. R. & Hatti-Kaul, R. Coenzyme a-acylating propionaldehyde dehydrogenase (pdup) from lactobacillus reuteri: Kinetic characterization and molecular modeling. *Enzyme and microbial technology* **53**, 235–242 (2013).

[79] Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega. *Molecular systems biology* **7**, 539 (2011).

[80] Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. The phyre2 web portal for protein modeling, prediction and analysis. *Nature protocols* **10**, 845–858 (2015).

[81] Kurcinski, M., Jamroz, M., Blaszczyk, M., Kolinski, A. & Kmiecik, S. Cabsdock web server for the flexible docking of peptides to proteins without prior knowledge of the binding site. *Nucleic acids research* **43**, W419–W424 (2015).

[82] Zhou, P., Jin, B., Li, H. & Huang, S.-Y. Hpepdock: a web server for blind peptide–protein docking based on a hierarchical algorithm. *Nucleic acids research* **46**, W443–W450 (2018).

[83] Lee, H., Heo, L., Lee, M. S. & Seok, C. Galaxypepdock: a protein–peptide docking tool based on interaction similarity and energy optimization. *Nucleic acids research* **43**, W431–W435 (2015).

[84] London, N., Raveh, B., Cohen, E., Fathi, G. & Schueler-Furman, O. Rosetta flexpepdock web server—high resolution modeling of peptide–protein interactions. *Nucleic acids research* **39**, W249–W253 (2011).

[85] Sutter, M., Greber, B., Aussignargues, C. & Kerfeld, C. A. Assembly principles and structure of a 6.5-MDa bacterial microcompartment shell. *Science* **356** (2017).

[86] Migneault, I., Dartiguenave, C., Bertrand, M. J. & Waldron, K. C. Glutaraldehyde: behavior in aqueous solution, reaction with proteins, and application to enzyme crosslinking. *Biotechniques* **37**, 790–802 (2004).

[87] Adams, P. D. *et al.* PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D: Biological Crystallography* **66**, 213–221 (2010).

[88] Krissinel, E. & Henrick, K. Protein interfaces, surfaces and assemblies service pisa at european bioinformatics institute. *J Mol Biol* **372**, 774–797 (2007).

[89] Carreira, E. M. & Yamamoto, H. *Comprehensive chirality* (Elsevier, 2012).

[90] Erbilgin, O., Sutter, M. & Kerfeld, C. A. The Structural Basis of Coenzyme A Recycling in a Bacterial Organelle. *PLoS Biology* **14**, 1–20 (2016).

[91] Dev, I. K. *et al.* Mode of binding of folate analogs to thymidylate synthase. evidence for two asymmetric but interactive substrate binding sites. *Journal of Biological Chemistry* **269**, 1873–1882 (1994).

[92] Hennig, M., Grimm, B., Contestabile, R., John, R. A. & Jansonius, J. N. Crystal structure of glutamate-1-semialdehyde aminomutase: an α2-dimeric vitamin b6-dependent enzyme with asymmetry in structure and active site reactivity. *Proceedings of the National Academy of Sciences* **94**, 4866–4871 (1997).

[93] Abeln, S. & Frenkel, D. Disordered flanks prevent peptide aggregation. *PLoS Comput Biol* **4**, e1000241 (2008).

[94] Aussignargues, C., Paasch, B. C., Gonzalez-Esquer, R., Erbilgin, O. & Kerfeld, C. A. Bacterial microcompartment assembly: the key role of encapsulation peptides. *Communicative & integrative biology* **8**, e1039755 (2015).

[95] Mayer, M. J. *et al.* Effect of bio-engineering on size, shape, composition and rigidity of bacterial microcompartments. *Scientific reports* **6**, 36899 (2016).

[96] De Vries, S. J., Van Dijk, M. & Bonvin, A. M. The haddock web server for data-driven biomolecular docking. *Nature protocols* **5**, 883 (2010).

[97] Dolinsky, T. J., Nielsen, J. E., McCammon, J. A. & Baker, N. A. Pdb2pqr: an automated pipeline for the setup of poisson–boltzmann electrostatics calculations. *Nucleic acids research* **32**, W665–W667 (2004).

[98] Fan, C. & Bobik, T. A. The n-terminal region of the medium subunit (pdud) packages adenosylcobalamin-dependent diol dehydratase (pducde) into the pdu microcompartment. *Journal of Bacteriology* **193**, 5623–5628 (2011).

[99] Sinha, S. *et al.* Alanine scanning mutagenesis identifies an asparagine–arginine–lysine triad essential to assembly of the shell of the pdu microcompartment. *Journal of molecular biology* **426**, 2328–2345 (2014).